



Universidade de Aveiro Secção Autónoma de Ciências da Saúde
2014

**PAULA MARIA
VAZ MARTINS**

**RESSONÂNCIA MAGNÉTICA EM ESTUDOS DE
PRODUÇÃO DE FALA**



**PAULA MARIA
VAZ MARTINS**

RESSONÂNCIA MAGNÉTICA EM ESTUDOS DE PRODUÇÃO DE FALA

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Ciências e Tecnologias da Saúde, realizada sob a orientação científica do Doutor António Joaquim da Silva Teixeira, Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro e do Doutor Augusto Marques Ferreira da Silva, Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro



Este trabalho foi realizado com o Apoio da FCT através da bolsa de doutoramento individual com a referência SFRH/BD/65183 /2009, financiada pelo Programa Operacional Potencial Humano - QREN Portugal 2007-2013 – Formação avançada, participado pelo Fundo Social Europeu e por fundos nacionais do MCTES.



dedicatória

Aos meus filhos Gonçalo e Maria Ana
À minha mãe
À memória do meu pai
Ao Rui Miguel

o júri

presidente

Doutor Luís Filipe Pinheiro de Castro

Professor Catedrático da Universidade de Aveiro

vogais

Doutor Luís Filipe Caseiro Alves

Professor Catedrático da Faculdade de Medicina da Universidade de Coimbra

Doutora Isabel Maria Martins Trancoso

Professora Catedrática do Instituto Superior Técnico da Universidade de Lisboa

Doutor Francisco António Cardoso Vaz

Professor Catedrático Aposentado da Universidade de Aveiro

Doutor Nelson Fernando Pacheco da Rocha

Professor Catedrático da Universidade de Aveiro

Doutor Miguel de Sá e Sousa de Castelo-Branco

Professor Associado da Faculdade de Medicina da Universidade de Coimbra

Doutor António Joaquim da Silva Teixeira

Professor Auxiliar da Universidade de Aveiro (Orientador)

Doutor Luís Jorge Oliveira Carrasco Lança

Professor Adjunto da Escola Superior da Tecnologia da Saúde de Lisboa

agradecimentos

Gostaria em primeiro lugar de agradecer aos meus orientadores pelos seus valiosos ensinamentos, por me criarem o gosto pela investigação, pela disponibilidade em me ajudar a superar os vários problemas que foram surgindo neste longo percurso. Agradeço-lhes também o facto de me terem pressionado, apenas, na justa medida do necessário para que este trabalho chegasse ao final. Ao Professor Augusto Silva fico agradecida pelo suporte na área do processamento de imagem e por me ajudar a encontrar alternativas simples, mas realistas, para resolver problemas complexos. A situação esteve frequentemente “desesperada mas não perdida”. Ao Professor António Teixeira agradeço, o ter-me feito descobrir esta área fantástica que é a produção de fala, também por tudo o que me ensinou e proporcionou durante estes últimos anos, que foi muito, mesmo muito, para além do seu trabalho de orientação científica. Obrigada pelas constantes palavras de incentivo e coragem, por estar sempre lá, sempre atento e pela compreensão que teve comigo nos momentos mais desesperados.

À Professora Catarina Oliveira agradeço todo o suporte que me deu na área da linguística, a infinita paciência para me explicar os termos e teorias linguísticas que demorei a perceber. Estou-lhe agradecida pelo rigor e exigência que colocou nas suas revisões, pelos seus comentários construtivos que em muito contribuíram para melhorar este trabalho. Para a Catarina são poucas as palavras para expressar o meu agradecimento por ter estado sempre ao meu lado, pelo incentivo e força que me transmitiu nos momentos mais complicados deste trabalho, pela excelente companhia e pelos momentos de boa disposição (peripécias incluídas) durante as nossas viagens de trabalho. É um privilégio poder ter-te como amiga - Obrigada Catarina!

No plano institucional, agradeço à FCT o apoio financeiro concedido através de uma Bolsa Individual de Doutoramento e o financiamento obtido no âmbito do projeto HERON II, imprescindível para a recolha dos dados de RM. Ao Professor Miguel Castelo-Branco, coordenador científico do IBILI, fico reconhecida por ter aberto as portas da instituição a este projeto; ao colega Carlos Ferreira agradeço sinceramente o seu empenho e o grande profissionalismo que revelou nas fases de otimização e aquisição de imagem e pela amizade que fomos estabelecendo. Agradeço à Professora Sónia Gonçalves a amabilidade e disponibilidade para o esclarecimento de algumas das minhas dúvidas.

À ESSUA, nomeadamente ao seu atual e anterior diretores, Professor Francisco Amado e Professor Nelson Rocha, agradeço o período de dispensa que me foi concedida; também aos colegas Sílvia De Francesco e Milton Santos que, pela minha ausência, ficaram mais sobrecarregados de trabalho, aqui fica o meu reconhecimento.

Ao IEETA agradeço o suporte logístico concedido e o excelente ambiente de trabalho proporcionado durante o período em que lá trabalhei. Aos seus funcionários, particularmente à Anabela, fico agradecida pela simpatia e prontidão com que sempre me ajudaram na resolução de assuntos burocráticos.

A todos os falantes que voluntariamente participaram neste estudo e que suportaram o tempo necessário dentro do túnel da RM aqui fica o meu agradecimento. Sem vocês não haveria sons, línguas ou tratos vocais para apresentar. Obrigada CO, MC, ER, SV, CM, JPM, JH, LCR, e AS pela vossa paciência.

Samuel, Liliana, Sara e Marisa agradeço sinceramente o vosso incentivo, a vossa ajuda sempre que dela necessitei e também a vossa insistência e preocupação para que eu finalizasse este trabalho. Ao Samuel agradeço a partilha de conhecimentos na área da visualização e computação gráfica e todo o seu empenho na segmentação das imagens em tempo real.

A toda a minha família, pais, irmão, cunhados, sobrinhos e sobrinhas, agradeço o apoio, a paciência, a compreensão que tiveram comigo durante este período. Aos Sampaio-Lopes e aos Grade-Silva por me conseguirem desviar do computador, dos *papers* . . . enfim, pelos momentos de agradável descontração e pela vossa amizade incondicional.

À minha mãe, agradeço a força que sempre me incutiu e por todos os sacrifícios que fez para que eu chegasse até aqui. Para o meu querido pai, um grande obrigada, que terá que chegar até ao Céu, por tudo o que foi e me transmitiu ao longo da sua vida.

Finalmente, uma palavra para os que mais perderam enquanto estive concentrada neste trabalho: o meu marido e os meus filhos Gonçalo e Maria Ana a quem dedico este trabalho. Ao Rui e ao Gonçalo um profundo reconhecimento por compreenderem as minhas ausências e por me concederem o espaço que necessitei para concluir esta longa empreitada. A ti MiAna, peço-te desculpa pelas muitas horas que estive contigo, não estando, pelas histórias que não te contei para contar esta que nunca mais terminava. Obrigada pelo teu sorriso, pelos teus beijinhos, que muito me alentaram, e pelo teu amor incondicional.

palavras-chave

Ressonância Magnética, Imagem volumétrica, Tempo Real, Produção de Fala, Português Europeu, Laterais, Vogais nasais.

resumo

Estudar os mecanismos subjacentes à produção de fala é uma tarefa complexa e exigente, requerendo a obtenção de dados mediante a utilização de variadas técnicas, onde se incluem algumas modalidades imagiológicas. De entre estas, a Ressonância Magnética (RM) tem ganho algum destaque, nos últimos anos, posicionando-se como uma das mais promissoras no domínio da produção de fala. Um importante contributo deste trabalho prende-se com a otimização e implementação de protocolos (RM) e proposta de estratégias de processamento de imagem ajustados aos requisitos da produção de fala, em geral, e às especificidades dos diferentes sons. Para além disso, motivados pela escassez de dados para o Português Europeu (PE), constitui-se como objetivo a obtenção de dados articulatórios que permitam complementar informação já existente e clarificar algumas questões relativas à produção dos sons do PE (nomeadamente, consoantes laterais e vogais nasais).

Assim, para as consoantes laterais foram obtidas imagens RM (2D e 3D), através de produções sustentadas, com recurso a uma sequência Eco de Gradiente (EG) rápida (3D VIBE), no plano sagital, englobando todo o trato vocal. O *corpus*, adquirido por sete falantes, contemplou diferentes posições silábicas e contextos vocálicos. Para as vogais nasais, foram adquiridas, em três falantes, imagens em tempo real com uma sequência EG - *Spoiled* (TurboFLASH), nos planos sagital e coronal, obtendo-se uma resolução temporal de 72 ms (14 frames/s). Foi efetuada aquisição sincronizada das imagens com o sinal acústico mediante utilização de um microfone ótico. Para o processamento e análise de imagem foram utilizados vários algoritmos semiautomáticos.

O tratamento e análise dos dados permitiu efetuar uma descrição articulatória das consoantes laterais, ancorada em dados qualitativos (e.g., visualizações 3D, comparação de contornos) e quantitativos que incluem áreas, funções de área do trato vocal, extensão e área das passagens laterais, avaliação de efeitos contextuais e posicionais, etc. No que respeita à velarização da lateral alveolar /l/, os resultados apontam para um /l/ velarizado independentemente da sua posição silábica. Relativamente ao /L/, em relação ao qual a informação disponível era escassa, foi possível verificar que a sua articulação é bastante mais anteriorizada do que tradicionalmente descrito e também mais extensa do que a da lateral alveolar. A resolução temporal de 72 ms conseguida com as aquisições de RM em tempo real, revelou-se adequada para o estudo das características dinâmicas das vogais nasais, nomeadamente, aspetos como a duração do gesto velar, gesto oral, coordenação entre gestos, etc. complementando e corroborando resultados, já existentes para o PE, obtidos com recurso a outras técnicas instrumentais. Para além disso, foram obtidos novos dados de produção relevantes para melhor compreensão da nasalidade (variação área nasal/oral no tempo, proporção nasal/oral).

Neste estudo, fica patente a versatilidade e potencial da RM para o estudo da produção de fala, com contributos claros e importantes para um melhor conhecimento da articulação do Português, para a evolução de modelos de síntese de voz, de base articulatória, e para aplicação futura em áreas mais clínicas (e.g., perturbações da fala).

keywords

Magnetic Resonance Imaging, Volumetric Imaging, Real time, European Portuguese, Speech Production, Lateral consonants, Nasal Vowels

abstract

The study of the mechanisms underlying speech production is a complex and demanding task that requires data gathered using different techniques and including image acquisition. Among the different imaging modalities used, Magnetic Resonance Imaging (MRI) assumed an important role, in recent years, positioning itself as one of the most promising techniques and providing a wealth of information concerning speech production. An important contribution of this research is the optimization and implementation of MRI protocols and the proposal of adequate image processing techniques that can meet the requirements imposed by speech production and the specificities of different sounds. Additionally, motivated by the scarcity of data for European Portuguese (EP), image acquisitions were performed to gather articulatory data to complement and clarify previous information relating to the production of EP sounds (namely, lateral consonants and nasal vowels).

For lateral consonants, MR images encompassing the entire vocal tract (VT), both in the midsagittal plane and in 3D, were acquired, during sustained productions, using a spoiled Gradient Echo (GE) sequence - 3D VIBE. The *corpus*, obtained for seven EP speakers, considered the lateral consonants in different syllabic contexts and syllable positions. For nasal vowels a *corpus* considering different syllabic positions and contexts was acquired, for three speakers, using Real-time MRI (RT- MRI) images by means of a GE - spoiled (TurboFLASH) sequence, obtained in the sagittal and coronal planes, with a temporal resolution of 72 ms (14 frames/s). A synchronized audio signal was acquired, inside the MR scanner using a fiberoptic microphone. Data processing and analysis was achieved using several semi-automatic algorithms.

Analysis of the acquired data allowed a detailed articulatory description of the lateral consonants anchored in both qualitative (e.g., 3D visualization, contour comparison) and quantitative data such as, vocal tract area functions, extension and area of lateral channels and evaluation of positional and contextual effects. Specifically, for the alveolar lateral /l/, as regards velarization, the gathered data points to a variety regardless of its syllabic position. For the /L/, in respect of which the information is very scarce, evidence shows the articulation is far more fronted than traditionally described and more extensive than that observed for the alveolar lateral.

The temporal resolution of 72 ms, achieved with RT- MRI acquisitions, proved to be suitable to address the study of dynamic characteristics of nasal vowels, namely velar and oral gestures, temporal coordination between gestures and durational aspects, complementing existing data for the EP, obtained using other instrumental techniques. In addition, new relevant data were attained providing additional contributions for a deep knowledge of nasality (e.g., nasal/oral areas over time, nasal/oral proportion).

The work presented demonstrates the versatility and potential of MRI when applied to speech production studies and provides important contributions to a better understanding of the articulation of EP, to the development of models supporting the improvement of articulatory based speech synthesis and to future applications in clinical areas (e.g., speech disorders).

Conteúdo

1	Introdução	1
1.1	Técnicas instrumentais para estudo e modelação da fala	2
1.2	Problemas e motivações	6
1.3	Objetivos	8
1.4	Resultados publicados	10
1.5	Estrutura da tese	13
2	Ressonância Magnética	15
2.1	Avanços tecnológicos na área da RM	17
2.1.1	Campo magnético principal - B_o	17
2.1.2	Gradientes de campo magnético	20
2.1.3	Tecnologia de antenas	23
2.1.4	Técnicas de aquisição rápida	25
2.1.4.1	Sequências de pulso rápidas e ultra-rápidas	25
2.1.4.2	Técnicas <i>Single-Shot</i> e Imagem Eco Planar	31
2.1.4.3	Trajetórias de preenchimento do Espaço K	33
2.1.4.4	Técnicas de subamostragem e outras técnicas para acelerar a aquisição	37
2.1.4.5	Imagem paralela	38
2.1.4.6	Técnicas emergentes	42
2.2	RM no âmbito de estudos de produção de fala	44
3	Processamento de Imagem	53
3.1	Segmentação de imagem	54

3.1.1	Técnicas de Segmentação de imagem	55
3.1.1.1	Nível de interação do utilizador	55
3.1.1.2	Outras classificações	57
3.1.2	Modelos deformáveis	57
3.1.2.1	<i>Snakes</i> - Modelos paramétricos	58
3.1.2.2	<i>Level-sets</i> - Modelos geométricos	59
3.1.2.3	<i>Live-wire</i> ou <i>Intelligent Scissors</i>	60
3.1.3	<i>Region Growing</i>	64
3.1.4	Ferramentas atualmente disponíveis	65
3.2	Segmentação de imagem nos estudos de produção de fala	66
3.2.1	Estudos envolvendo aquisições 2D e 3D estáticas	66
3.2.2	Estudos dinâmicos	70
3.3	Os problemas de processamento de imagem nesta área	74
3.4	Métodos de segmentação e ferramentas utilizadas neste estudo	76
3.4.1	ITK-SNAP e MeVisLab	78
4	Consoantes Laterais e Vogais Nasais	83
4.1	As consoantes laterais	85
4.1.1	A lateral alveolar /l/	85
4.1.1.1	Características articulatórias	85
4.1.1.2	Características acústicas	91
4.1.1.3	O /l/ do Português Europeu	92
4.1.2	A lateral palatal - /L/	94
4.1.2.1	Características articulatórias	94
4.1.2.2	Características acústicas	96
4.1.2.3	O /L/ no Português Europeu	97
4.2	Nasalidade	99
4.2.1	O véu palatino e a passagem velofaríngea	99
4.2.2	Grau de abertura, altura do véu palatino	100
4.3	Vogais nasais	101
4.3.1	Breves notas sobre a evolução histórica das vogais nasais no Português e seu estatuto fonológico	102
4.3.2	Aspetos dinâmicos e temporais da nasalidade vocálica no PE	103
4.3.2.1	As três fases da vogal nasal	104

4.3.2.2	Velocidade, duração dos gestos de abertura e fecho do véu palatino e coordenação entre gestos	105
4.3.3	A configuração oral das vogais nasais	106
5	RM Aplicada ao Estudo das Consoantes Laterais do Português	109
5.1	Estudos RM para estudo e modelação das consoantes laterais	110
5.2	Caracterização do problema e questões a investigar	112
5.2.1	Objetivos do estudo	112
5.3	Método	113
5.3.1	<i>Corpus</i>	113
5.3.2	Falantes/Informantes	114
5.3.3	Recolha de dados	116
5.3.3.1	Local e Período	116
5.3.3.2	Equipamento e material	116
5.3.3.3	Sessões preparatórias	116
5.3.3.4	Protocolo de aquisição de RM	119
5.3.3.5	Recolha de dados acústicos	121
5.3.4	Processamento de Imagem	122
5.3.4.1	Segmentação das arcadas e trato vocal	123
5.3.4.2	Segmentação da língua	127
5.3.4.3	Outras experiências de segmentação efetuadas	134
5.4	Avaliação (preliminar) das segmentações efetuadas para a língua com <i>ITK-SNAP</i> .	138
5.4.1	Avaliação qualitativa	140
5.4.2	Avaliação quantitativa	142
5.4.2.1	Análise do volume total da língua - um falante	143
5.4.2.2	Análise do volume total da língua - vários falantes	144
5.4.2.3	Métricas de semelhança (<i>Jaccard e Voxelwise difference</i>) - informação espacial	145
5.4.2.4	Discussão dos resultados relativos ao estudo de avaliação	147
5.5	Resultados	148
5.5.1	A lateral alveolar - /l/	149
5.5.1.1	Ponto e extensão do contacto alveolar na linha média	149
5.5.1.2	Configuração da língua em 2D e 3D	151
5.5.1.3	Funções de área do trato vocal	155

5.5.1.4	Passagens laterais	161
5.5.1.5	Efeito da posição silábica (2D e 3D)	168
5.5.1.6	Efeito do contexto vocálico	174
5.5.2	A lateral palatal - /L/	177
5.5.2.1	Ponto e extensão do contacto linguo-palatal	177
5.5.2.2	Configuração da língua 2D e 3D	179
5.5.2.3	Funções de área do trato vocal e passagens laterais	180
5.5.2.4	Efeito do contexto vocálico	184
5.5.2.5	Utilização da informação numa experiência de simulação para o /L/	186
5.5.3	Comprimento do trato vocal	191
5.6	Discussão	191
6	RM em Tempo Real para o Estudo das Vogais Nasais do Português	199
6.1	RM em tempo real	199
6.2	RM tempo real aplicada ao estudo da nasalidade vocálica	202
6.3	Justificação e objetivos do estudo	208
6.4	Método	209
6.4.1	<i>Corpus</i>	209
6.4.2	Falantes	210
6.4.3	Recolha de dados	211
6.4.3.1	Local e Período	211
6.4.3.2	Equipamento e material	211
6.4.3.3	Sessões Experimentais	212
6.4.3.4	Protocolo de aquisição final	212
6.4.4	Processamento do sinal de voz	216
6.4.5	Processamento de Imagem	216
6.4.5.1	Extração de contornos	217
6.4.5.2	Extração de informação relativa ao movimento e coordenação dos articuladores no tempo	219
6.4.5.3	Obtenção de áreas (nasais/orais) na zona de acoplamento	223
6.5	Resultados	226
6.5.1	Caracterização dos movimentos do véu palatino	226
6.5.1.1	Duração dos movimentos abertura-fecho do véu palatino	231
6.5.2	Articulação oral das vogais nasais	232

6.5.3	Áreas nasais e orais na zona de acoplamento	239
6.5.3.1	Variação temporal	239
6.5.3.2	Área máxima de abertura do PVF	240
6.5.3.3	Proporção nasal/oral (P_{NO})	242
6.5.4	Organização temporal dos movimentos dos articuladores durante a produção de vogais nasais	243
6.6	Discussão	247
6.6.1	Aspetos técnicos	247
6.6.2	Articulação das vogais nasais	250
6.6.2.1	Movimentos do véu palatino	250
6.6.2.2	Articulação oral das vogais nasais	251
6.6.2.3	Variação das áreas no tempo	252
6.6.2.4	Aspetos de coordenação temporal/produção da vogal nasal em três fases	254
7	Conclusões	257
7.1	Resumo do trabalho e principais contributos	257
7.2	Trabalho Futuro	261
7.3	Epílogo	264
	Bibliografia	265
A	Informação ao Participante	305
B	Registo para Caracterização do Participante	309
C	Declaração de Consentimento Informado	311
D	Questionário Rastreio de Ressonância Magnética	313
E	Artigo - IJCICG	315
F	Otimização de Protocolos e Procedimentos	333
F.1	Escolha e Otimização de sequências de pulso	333
F.1.1	Algumas considerações acerca da sequência TurboFLASH	339
F.2	Testes com Microfone ótico (MO), sincronização de imagem/som e codificação . . .	342
G	Corpus Nasais	345

G.1	Aquisições em Tempo Real - Nasais	345
G.1.1	Vogal nasal em posição inicial, medial e final - Contexto de oclusiva bilabial [p]	345
G.1.2	Vogal nasal em posição inicial, medial e final - Contexto de fricativa [f] . . .	345
G.1.3	Vogais orais produzidas em contexto de oclusiva bilabial [p]	345
G.1.4	Sequência de vogais orais e nasais produzidas de forma isolada	346
G.1.5	Oblíquos e duplos oblíquos	346
G.1.6	Consoantes nasais - posição intervocálica	346
G.1.7	Aquisição com frase suporte - Diz XXX após	346
G.2	Material Complementar - Laterais e róticos	346

Lista de Figuras

2.1	Atuação dos gradientes de campo magnético. Fonte da imagem: Brown e Semelka (2010).	20
2.2	Representação esquemática de uma bobine <i>phased array</i> de 4 canais (à esquerda). Cada bobine tem o seu perfil de sensibilidade relativo a B_1 que é depois conjugado com o de outras bobines. Fonte da imagem: Fujita <i>et al.</i> (2006).	24
2.3	Antenas <i>phased-array</i> para cobertura anatômica de todo o corpo com elevado número de canais. Fonte da imagem: Brix <i>et al.</i> (2008).	24
2.4	Diagrama de pulsos de uma sequência EG - <i>Spoiled</i> . Fonte da imagem: Chavhan <i>et al.</i> (2008).	27
2.5	Diagrama de pulsos de uma sequência EG b-SSFP. Fonte da imagem: Chavhan <i>et al.</i> (2008).	30
2.6	Técnica <i>Half-Fourier</i> . Fonte da imagem: Brix <i>et al.</i> (2008).	32
2.7	Modo de aquisição com esquema de leitura eco-planar. À esquerda EG-EPI e à direita SE-EPI. Fonte das imagens: Brix <i>et al.</i> (2008).	33
2.8	Relação entre o padrão de amostragem no domínio da frequência espacial (espaço K) e o resultado no domínio da imagem. Fonte da imagem: Lustig <i>et al.</i> (2008).	34
2.9	Exemplos de trajetórias de preenchimento do espaço K. Fonte da imagem: Lustig <i>et al.</i> (2008).	35
2.10	Técnica <i>View Sharing</i> com reconstrução por <i>sliding windows</i> . Fonte da imagem: Wieben (2007)	38
2.11	Reconstrução da imagem com técnica GRAPPA. Fonte da imagem: Deshmane <i>et al.</i> (2012).	40
2.12	Exemplo de imagens adquiridas com técnica GRAPPA. Fonte da imagem: Deshmane <i>et al.</i> (2012).	42

2.13	Exemplo de aquisição 2D em vários planos (axial, oblíquo e coronal) para obtenção de informação 3D da língua. Fonte da imagem: Badin <i>et al.</i> (2002).	46
3.1	Tarefa de co-registo (manual) das arcadas dentárias de um dos falantes participantes no estudo (LCR) com o volume adquirido durante a produção sustentada do /l/ para viabilizar a identificação das passagens laterais.	54
3.2	Taxonomia relativa aos modelos deformáveis.	58
3.3	Diagrama que representa um grafo. Fonte da imagem: Picado (2010).	61
3.4	Segmentação do trato vocal em imagens sagitais, utilizando <i>live-wire</i> ao longo de uma série de cortes 2D pertencentes a um volume de dados adquirido no âmbito deste trabalho.	62
3.5	<i>Live-wire 3D</i> para segmentação de objetos com topologias complexas. Fonte da imagem: Poon <i>et al.</i> (2008).	63
3.6	Visualizações 3D da língua obtidas com <i>Region Growing</i> . Fonte da imagem: Bone <i>et al.</i> (2011).	68
3.7	Utilização de <i>Live-wire 3D</i> para segmentação do véu palatino. Fonte da imagem: Fels <i>et al.</i> (2009).	68
3.8	Segmentação com contornos geodésicos ativos refinada com recurso a <i>level-set</i> laplaciano. Imagem nos três planos referente à segmentação da língua para um som rótico. Fonte de imagem: Vogt <i>et al.</i> (2009).	69
3.9	Segmentação com algoritmo de <i>Random Walker</i> . Visualização 3D da língua obtida a partir de segmentação manual e com o algoritmo semi-automático (em baixo, à direita). Fonte da imagem: Suo <i>et al.</i> (2011).	70
3.10	Exemplo de estudo em tempo real com segmentação manual do trato vocal. Fonte da imagem: Kedrova <i>et al.</i> (2009).	70
3.11	Segmentação de imagens 2D em tempo real com <i>kalman snakes</i> e <i>optical flow</i> . Fonte de imagem: Narayanan <i>et al.</i> (2004).	72
3.12	Segmentação no domínio da frequência. Fonte da imagem: Bresch e Narayanan (2009).	72
3.13	Modelo geométrico do trato vocal (esquerda) e descritores do trato vocal (direita). Fonte da imagem: Bresch e Narayanan (2009) e Ramanarayanan <i>et al.</i> (2010). . . .	73
3.14	Identificação da zona e extensão da constrição na produção de sons consonânticos. O exemplo refere-se às consoantes coronais [t, n, l]. Fonte da imagem: Hagedorn <i>et al.</i> (2011).	73
3.15	Exemplos de imagens, adquiridas no âmbito deste trabalho, apresentando informação obtida de forma estática (2D e 3D) e dinâmica (<i>2D+t</i>) durante a produção de fala.	75

3.16	Exemplo ilustrativo de imagens do trato vocal adquiridas, neste estudo, durante a produção sustida de fonemas.	76
3.17	Exemplo de imagens adquiridas em tempo real, para um dos falantes (CO) deste estudo, durante a produção de fala com uma resolução temporal de 72 <i>ms</i> (14 <i>frames/s</i>).	77
3.18	Interface gráfica (GUI) do <i>MeVisLab</i> . Rede desenvolvida para efetuar MPR curvo. Exemplo efetuado para um dos falantes participantes neste estudo, relativo ao som [L].	80
3.19	Exemplo de segmentação do lobo caudado com os dois algoritmos disponibilizados pelo <i>ITK-SNAP</i> . Fonte da imagem: Yushkevich <i>et al.</i> (2006).	80
4.1	Configurações do trato vocal para o /l/ velarizado (<i>dark-l</i>), à esquerda, e não velarizado (<i>clear-l</i>), à direita. Fonte da imagem:Recasens (2012).	87
4.2	Configuração articulatória 2D do /l/ em posição inicial e posição final para um falante do PE. Fonte da imagem: Martins (2007).	94
4.3	Padrão de contacto da lateral alveolar (à esquerda) e da lateral palatal (à direita) num falante do Português Brasileiro. Fonte da imagem: Reis e Espesser (2006).	97
4.4	Músculos responsáveis pelos movimentos do véu palatino e da passagem velofaríngea. Tensor do véu palatino (1), Elevador do véu palatino (2), Palatoglossos (3), Palatofaríngeo (4) e Constritor da faringe (5). Fonte da imagem: Serrurier e Badin (2008).	100
5.1	Sequência de passos para efetuar a aquisição das imagens. Exemplo para a eliciação do /l/, em início palavra, como em laca	120
5.2	Exemplo de segmentação das arcadas dentárias do falante JPM - mandíbula (esquerda), maxilar (centro) e malha 3D da mandíbula (direita).	124
5.3	Rede <i>MeVisLab</i> criada para efetuar o co-registo das arcadas dentárias dos falantes com os volumes obtidos durante a produção de fonemas.	124
5.4	Rede <i>MeVisLab</i> utilizada para efetuar Reconstrução Multiplanar Curva ao longo do trato vocal. Exemplo para o falante LCR na produção de /l/ como em sal	125
5.5	Rede <i>MeVisLab</i> , com módulo <i>Live-wire</i> , para segmentação das imagens 2D da língua.127	
5.6	Rede <i>MeVisLab</i> , com módulo <i>Live-wire</i> , para segmentação e visualização das imagens 3D da língua.	129
5.7	Interface gráfica do <i>ITK-SNAP</i> . <i>Display</i> do volume nos planos axial e sagital (em cima) e coronal (em baixo). Exemplo de segmentação numa produção do falante JPM. Malha superficial do véu palatino (em baixo, à esquerda).	130
5.8	Seleção da região de interesse e re-amostragem dos dados.	130
5.9	Limiares selecionados para criar mapa de probabilidades.	132

5.10	Seleção dos parâmetros para os termos da equação que governam o processo de evolução da <i>snake</i>	132
5.11	Colocação das bolhas (<i>balloons</i>) dentro da região de interesse e evolução iterativa da segmentação.	133
5.12	Revisão das segmentações e correções manuais efetuadas pelo operador.	133
5.13	Influência dos algoritmos de suavização e decimação nas malhas superficiais.	134
5.14	Dados estatísticos relativos ao volume, fornecidos pelo <i>ITK-SNAP</i>	134
5.15	Exemplo de malha 3D resultante de segmentação do trato vocal para uma vogal, efetuada com algoritmo semi-automático no <i>ITK-SNAP</i> . Falante MC - vogal [u]. . .	135
5.16	Malhas 3D da língua e trato vocal obtidas com o <i>ITK-SNAP</i> , para 3 falantes. . . .	136
5.17	Exemplo de segmentação das arcadas dentárias e véu palatino da falante MC. À esquerda imagem da mandíbula e à direita imagem da maxila e véu palatino. . . .	136
5.18	Exemplo de segmentação da faringe para uma vogal com recurso ao algoritmo de <i>Region Growing</i> 3D através do <i>MeVisLab</i>	137
5.19	Exemplo de segmentação manual, guiada nos três planos, efetuada com o <i>ITK-SNAP</i> . Imagem publicada em Martins <i>et al.</i> (2011c).	140
5.20	Sobreposição de malhas em 3D, para comparar as diferentes segmentações SA efetuadas. Exemplo para três falantes com produções de diferentes sons ([l] para os falantes JPM e CO e [L] para o falante LCR).	142
5.21	Áreas notáveis consideradas na análise de semelhança entre as regiões X e Y. As fórmulas de cálculo para o coeficiente de <i>Jaccard</i> e <i>Voxelwise difference</i> são apresentadas na parte inferior da figura. Imagem adaptada de Silva <i>et al.</i> (2011) com permissão dos autores.	146
5.22	<i>Boxplots</i> para as métricas de sobreposição, para todos os pares comparados. À esquerda, coeficiente de <i>Jaccard</i> [0,1]; <i>voxelwise difference</i> (%), à direita.	146
5.23	Representação do trato vocal com as zonas e sub-zonas articulatórias e regiões da língua: (1) zona alveolar, (2) pré-palatal, (3) médio-palatal, (4) pós-palatal, (5) ápice e lâmina lingual, (6) pré-dorso, (7) parte média do dorso e (8) pós-dorso. . .	149
5.24	Imagens 2D no plano médio-sagital relativas a seis falantes durante a produção do /l/ em posição de ataque (esquerda) e em coda (direita). Ponto de articulação da lateral e configuração da língua e trato vocal em 2D. De cima para baixo: falantes CO, MC, ER, LCR, JH e JPM.	150
5.25	Imagens da língua no plano coronal durante a produção da lateral em posição de ataque [lak6] no contexto de [a] (à esquerda) e na produção da vogal [a] (à direita). . .	153
5.26	Malhas 3D da língua obtidas com o <i>ITK-SNAP</i> . Em cima, /l/ produzido, pela falante CO, em ataque e, em baixo, em posição de coda.	154

5.27	Malha superficial da língua onde se observa a presença de <i>groove</i> ao longo da linha média. Imagem referente ao falante JPM na produção do /l/, em ataque.	155
5.28	Áreas segmentadas ao longo do trato vocal para o falante JPM, na produção de [lak6].	156
5.29	Áreas segmentadas ao longo do trato vocal para a falante CO, na produção de [lak6].	157
5.30	Áreas segmentadas ao longo do trato vocal para o falante LCR, na produção de [lit4u].	158
5.31	Função de área anotada referente ao /l/ para o falante JPM, no contexto da vogal [i], em posição de ataque intervocálico [biliS].	159
5.32	Exemplos de funções de área referentes ao /l/ para quatro falantes (CO, MC, JPM e LCR) na produção de [lak6].	160
5.33	Função de área normalizada, para seis falantes, referente ao /l/ na produção de [lak6].	161
5.34	Funções de área normalizadas (para seis falantes) referentes ao /l/ em vários contextos e posições.	162
5.35	Imagens em diferentes zonas do trato vocal, do falante JPM, que mostram o fluxo de ar lateral, as passagens laterais, a cavidade supra-lingual e as zonas de contacto da língua com o palato lateralmente e na linha média.	163
5.36	Exemplo de visualização 3D do trato vocal - série de cortes empilhados obtidos para o /l/ em posição inicial como em [lak6] para os falantes CO (em cima) e LCR (em baixo). É possível observar a zona de contacto alveolar, as passagens laterais e a cavidade supra-lingual separada (para o falante LCR). As restantes zonas do trato vocal estão identificadas.	164
5.37	Visualizações 3D do trato vocal, ilustrando as passagens laterais, para dois falantes (CO e LCR), para o /l/ produzido em posição inicial e final, no contexto das vogais [a], [i] e [u]	167
5.38	Sobreposição dos contornos 2D para seis falantes para o /l/, em posição de ataque (LV), ataque intervocálico (VLV) e coda (VL), no contexto da vogal [a]. Falantes femininas CO, MC, ER (em cima) e, em baixo, falantes masculinos (AS, JPM, LCR).	168
5.39	Efeito da posição silábica - Sobreposição dos contornos 2D para uma falante (CO) considerando produções do /l/ em posição de ataque (LV), ataque intervocálico (VLV) e coda (VL), no contexto das vogais [a] (esquerda), [i] (centro), [u] (direita).	169
5.40	Média dos IP obtidos para todos os falantes, nas diferentes posições silábicas (esquerda). IP médios para o par ataque (pi) e coda (pf), em função da vogal em que foi produzido (direita).	170
5.41	Média dos IP por falante (à esquerda) e por género (à direita), considerando todas as posições silábicas.	170

5.42	Funções de área do trato vocal para o /l/, em posição de ataque ([lak6]) e de coda ([sal]), para cada vogal [a, i, u] referentes à falante CO.	171
5.43	Funções de área do trato vocal para o /l/, em posição de ataque (vermelho) e de coda (verde), no contexto da vogal [a] referentes aos falantes CO, MC, LCR e JPM. 171	
5.44	Funções de área do trato vocal para o /l/, em posição de ataque (vermelho) e de coda (verde), no contexto da vogal [i] referentes aos falantes CO, MC, LCR e JPM. 171	
5.45	Funções de área do trato vocal para o /l/, em posição de ataque (vermelho) e de coda (verde), no contexto da vogal [u] referentes aos falantes CO, MC, LCR e JPM. 172	
5.46	Exemplo de função de área normalizada, para seis falantes, na produção de [lak6], em cima e [sal] em baixo. A curva representa o valor médio da área. As diferentes zonas do trato são identificadas na imagem referente a [lak6].	172
5.47	Exemplo de função de área normalizada, para seis falantes, na produção de [lup6], em cima e [sul] em baixo. A curva representa o valor médio da área. As diferentes zonas do trato são identificadas na imagem referente a [lup6].	173
5.48	Variabilidade entre falantes, na produção do /l/ em ataque e coda, observada através da sobreposição das malhas 3D da língua (à esquerda) e respectivas funções de área do trato vocal (à direita), para os falantes JPM (em cima) e CO (em baixo). . . .	174
5.49	Sobreposição dos contornos 2D da língua, extraídos das imagens no plano médio sagital, nas produções referentes ao /l/, em posição de ataque, no contexto das vogais [a, i, u].	174
5.50	Efeito do contexto - Funções de área do trato vocal para o /l/, produzido no contexto das vogais [a] [i] e [u] nas três posições silábicas consideradas para dois falantes. . .	175
5.51	Efeito do contexto - Funções de área do trato vocal para o /l/, produzido no contexto das vogais [a] [i] e [u], em posição inicial para 4 falantes.	176
5.52	Efeito do contexto - Funções de área do trato vocal para o /l/, produzido no contexto das vogais [a] [i] e [u], em posição final para 4 falantes.	176
5.53	Efeito contextual em 3D - sobreposição e comparação de malhas 3D da língua, efetuada com recurso ao PolyMeco, para avaliar efeitos coarticulatórios na produção do /l/ em ataque (à esquerda) e em coda (à direita) para a falante CO. Imagem publicada em Martins <i>et al.</i> (2012a).	177
5.54	Ponto de articulação do [L] para seis falantes.	178
5.55	Ponto de articulação do [L] para a falante ER.	179
5.56	Visualizações 3D da língua para o /l/ (em cima) e para o /L/ (em baixo), relativas à falante CO, no contexto da vogal [a]. Da esquerda para a direita: vista anterior, lateral, posterior e superior.	180
5.57	Áreas segmentadas ao longo do trato vocal para o falante LCR, na produção de [paL6].	181

5.58	Áreas segmentadas ao longo do trato vocal para a falante MC, na produção de [paL6].	182
5.59	Funções de área do trato vocal para 2 falantes femininos (CO e MC) e dois masculinos (JPM e LCR) relativos ao /L/ produzido em [paL6].	183
5.60	Sobreposição das funções de área do trato vocal na produção do [l] e [L] em posição intervocálica no contexto da vogal [i]. Exemplo para a falante CO e falante LCR. .	184
5.61	Efeito contextual em 2D - Sobreposição dos contornos 2D obtidos durante a produção do [L] no contexto das vogais cardinais para duas falantes femininas (CO e MC) e para os quatro falantes do gênero masculino (AS, JH, JPM e LCR).	185
5.62	Funções de área do trato vocal para 2 falantes femininos (CO e MC) e dois masculinos (JPM e JH) relativos ao /L/ produzido no contexto das três vogais como em [paL6], [piL6] e [puL6].	186
5.63	Representação em esquema das diferentes cavidades consideradas para a simulação e visualizações 3D da língua sobre um volume com as arcadas dentárias registradas que mostram as cavidades posterior, supra-lingual e frontal e as passagens laterais. Imagens do falante LCR durante a produção do [L] em palha.	187
5.64	Função de resposta acústica do trato vocal obtida a partir de simulações efetuadas em diferentes condições: dois canais laterais e cavidade supra-lingual (preto), dois canais laterais sem cavidade supra-lingual (vermelho) e apenas um canal lateral sem cavidade supra-lingual (verde). Valores da ordenada em <i>dB</i> e abcissa em <i>Hz</i>	189
5.65	Tipo de informação retirado dos volumes de dados de RM	193
6.1	Posicionamento do falante e colocação do MO na bobine de crânio	213
6.2	Em cima, orientação escolhida para a marcação dos cortes coronais nos lábios (linha amarela) e no véu palatino (linha vermelha). Em baixo, exemplo das imagens obtidas: resolução temporal de 144 <i>ms</i>	215
6.3	Exemplo ilustrativo da forma como se poderá obter informação sincronizada para os dois articuladores, através de um processo de interpolação linear.	215
6.4	Sinal acústico, espetrograma e respetiva anotação fonética para a sequência ampa , pampa , pan para uma das produções da falante CO	217
6.5	Exemplo de segmentações que necessitam de edição e correção por parte do operador	218
6.6	Sobreposição de contornos relativos à mesma vogal para cada uma das três repetições efetuadas com vogal nasal em posição medial no contexto de oclusiva bilabial. Falante CM - [6~] (à esquerda), falante SV - [e~] (ao centro) e falante CO - [o~] (à direita).	219
6.7	Exemplo de sobreposição de contornos 2D, para a falante CO. À esquerda representados três contornos ao longo da produção de uma vogal nasal [e~] e à direita para a vogal oral [6].	219
6.8	Passos referentes à aplicação do método baseado na intensidade média dos pixels .	221

6.9	Curvas de variação do véu palatino e dos lábios obtidas com o método baseado na intensidade média do <i>pixel</i> . Sequência com uma das vogais nasais.	222
6.10	Posição do véu palatino (aberto/fechado) e abertura labial obtida a partir dos <i>pixels</i> ativos dentro de uma ROI. Imagem publicada em Silva <i>et al.</i> (2013a).	222
6.11	Exemplo das curvas obtidas que permitem inferir os movimentos do véu palatino e dos lábios.	223
6.12	Orientação dos cortes coronais no véu palatino (à esquerda), identificação das cavidades nasal e oral (centro) e a série de imagens obtidas ao longo do tempo (direita).	223
6.13	Interface gráfica implementada com o <i>MeVisLab</i> , para segmentação das cavidades oral e nasal.	224
6.14	Em cima, curva de variação da área nasal com os segmentos correspondentes às diferentes fases assinalados com diferentes cores. Em baixo, primeira derivada da curva de área nasal com as referências utilizadas para determinar a duração do gesto de abertura (<i>Dopening</i>), fase estável (<i>plateau</i>) e fecho do véu palatino (<i>Dclosing</i>).	225
6.15	Cortes sagitais durante a produção do [6~], em posição inicial, como em ampa [6~p6] produzido pela falante CO.	227
6.16	Cortes sagitais durante a produção do [6~], em posição medial, como em pampa [p6~p6] produzido pela falante CO.	228
6.17	Cortes sagitais durante a produção do [6~], em posição final, como em pan [p6~] produzido pela falante CO.	229
6.18	Exemplo de curvas que permitem acompanhar os movimentos do véu palatino durante a sequência com as vogais nasais em diferentes posições. Em cima, sequência com a vogal [6~], ao centro o [i~] e [o~] em baixo.	230
6.19	Valores médios da duração (em <i>ms</i>) dos movimentos de abertura, zona estável e fecho do véu palatino para as falantes CO e CM.	232
6.20	Valores médios da duração (em <i>ms</i>) dos movimentos de abertura, zona estável e fecho do véu palatino para as falantes CO e CM, para cada uma das cinco vogais nasais do PE [6~],[e~],[i~],[o~] e [u~].	232
6.21	Comparação da articulação oral/nasal efetuada através da sobreposição de contornos relativos aos pares [i]/[i~] (em cima) e [u]/[u~] (em baixo) para as falantes CO (esquerda) e SV (direita). Imagem publicada em Oliveira <i>et al.</i> (2012).	234
6.22	Comparação da articulação oral/nasal através da sobreposição dos contornos relativos às vogais [a]/[6]/[6~] (em cima) e [O]/[o]/[o~] (em baixo) para as falantes SV (esquerda) e CM (direita). Imagem publicada em Oliveira <i>et al.</i> (2012).	235
6.23	Sobreposição de <i>frames</i> obtidas em diferentes momentos temporais ao longo da produção das vogais nasais [6~] (em cima), [e~] (ao centro) e [i~] (em baixo) produzidas pela falante CO. Da esquerda para a direita - produção de cada vogal em posição inicial, medial e final.	237

6.24	Sobreposição de <i>frames</i> obtidas em diferentes momentos temporais ao longo da produção das vogais nasais [o~] (em cima), e [u~] (em baixo) produzidas pela falante CO. Da esquerda para a direita - produção de cada vogal em posição inicial, medial e final.	238
6.25	Sobreposição dos contornos obtidos ao longo do tempo (<i>frames</i> relativas ao início, meio e final) durante a produção da vogal [o~] produzida de forma isolada (esquerda) e em posição medial em contexto de [p] como em [po~p6]. Exemplo ilustrativo para a falante CM.	238
6.26	Sobreposição dos contornos obtidos ao longo do tempo (<i>frames</i> relativas ao início, meio e final) durante a articulação da vogal [u~] produzida de forma isolada pela falante CO (esquerda) e pela falante SV (direita).	239
6.27	Exemplo da variação das áreas orais e nasais ao longo do tempo (falante CO) para todas as vogais nasais (da esquerda para a direita - [6~], [e~], [i~], [o~], [u~]) produzidas em início de palavra (em cima), posição medial (ao centro) e posição final (em baixo).	240
6.28	Área máxima nasal ou abertura do PVF (mm^2) em função da vogal nasal (à esquerda) e posição silábica (à direita) para as 3 falantes.	241
6.29	Curvas de variação do véu palatino e lábios, ao longo do tempo, para todas as vogais nasais, efetuadas pela falante CO.	244
6.30	Medidas de sincronização entre o gesto do véu palatino e o gesto oral. Falante CO durante a produção de [o~] como produzido na palavra [o~p6].	245
6.31	Exemplo de coordenação sequencial entre o gesto de abertura do véu palatino e o gesto de fecho labial, durante a produção de [i~p6].	246
6.32	Exemplo de coordenação sequencial entre o gesto de abertura do véu palatino e gesto de fecho labial, durante a produção de [po~p6].	246
6.33	Identificação das três fases na produção das vogais nasais. Falante CO durante a produção de [6~] como produzido na palavra [6~p6].	247
6.34	Identificação das diferentes fases na produção das vogais nasais. Falante CO durante a produção de [6 ~] como produzido na palavra [p6 ~].	248
7.1	Resumo dos contributos da tese nas diferentes áreas do trabalho e o seu potencial de aplicação.	258
7.2	Exemplo de imagens que podem ser utilizadas para o ensino da fonética. Imagem publicada em Martins <i>et al.</i> (2012a)	263
F.1	Exemplo de artefactos de <i>aliasing</i> devido à escolha de FOV demasiado pequeno ao longo da direcção de codificação de fase (sentido A-P).	335

F.2	Exemplos de algumas imagens (adquiridas com diferente número de <i>frames/s</i>), obtidas em fantoma, com sequências SS-TurboFLASH, SS-TSE (HASTE), TrueFISP e SS-TrueFISP. Artefactos de suscetibilidade magnética (<i>Banding artefacts</i>) mais evidentes com a sequência TrueFISP.	336
F.3	Exemplos de imagens obtidas com algumas sequências EG, testadas no âmbito deste estudo.	337
F.4	Perda de sinal observada ao longo da sequência de imagens devido a efeitos de saturação. Em cima, sequência FISP e em baixo, a sequência utilizada - TurboFLASH.	337
F.5	Imagens adquiridas com a sequência TurboFLASH - Efeito da variação do <i>Flip Angle</i> e impacto no SAR.	338
F.6	Exemplo de imagens obtidas com a sequência otimizada, com diferente resolução temporal.	338
F.7	Diagrama de pulsos da sequência TurboFLASH efetuada com o programa de simulação POET.	341
F.8	<i>Output</i> do Microfone ótico.	342

Lista de Tabelas

2.1	<i>Performance dos gradientes - Implicações práticas. Fonte da imagem: Lafuente (2000).</i>	22
2.2	<i>Exemplos de estudos estáticos com RM.</i>	47
2.3	<i>Exemplos de estudos com RM em tempo real</i>	48
2.4	<i>Exemplos de estudos realizados com RM para várias línguas.</i>	51
2.5	<i>Estudos de RM na produção e modelação fala - áreas de aplicação.</i>	52
5.1	<i>Corpus de laterais adquirido com RM (2D e 3D)</i>	114
5.2	<i>Parâmetros de aquisição utilizados na sequência TSE T1 marcada, na linha média, no plano sagital.</i>	121
5.3	<i>Parâmetros de aquisição da sequência EG 3D FLASH VIBE otimizada para aquisição de imagens durante elocução dos diferentes fonemas (3D Fonação) e para as arcadas dentárias dos falantes (3D Arcadas).</i>	122
5.4	<i>Resumo do trabalho de processamento efetuado.</i>	138
5.5	<i>Volume da língua (mm^3) obtido para 14 produções diferentes da falante CO.</i>	143
5.6	<i>Volume da língua (mm^3) obtido para vários falantes. Valor médio e desvio padrão (DP) para os métodos manual (M) e semi-automático (SA).</i>	144
5.7	<i>Resultados obtidos após cálculo do coeficiente de correlação intra-classes (CCI).</i>	145
5.8	<i>Mínimo, Máximo, Média e DP obtidos para as duas métricas a partir de todos os pares comparados. O coeficiente de Jaccard varia entre $[0, 1]$ e o Voxelwise difference é apresentado em (%).</i>	147
5.9	<i>Resultados ANOVA relativos às métricas (Jaccard e Voxelwise difference), considerando o método de segmentação como fator.</i>	147
5.10	<i>Extensão dos contactos na linha média para todos os falantes. Valores para o /l/ em ataque, ataque intervocálico (At.inter) e coda.</i>	152

5.11	<i>Extensão das passagens laterais (mm) e área média das passagens laterais (cm²). Informação para duas falantes (CO e MC) e dois falantes (JPM e LCR).</i>	165
5.12	<i>Extensão das passagens laterais em (mm), valor médio, máximo e mínimo, por falante.</i>	165
5.13	<i>Extensão das passagens laterais (mm), valor médio para as passagens esquerda e direita, em função da posição silábica.</i>	166
5.14	<i>Extensão das passagens laterais em (mm), valor médio para as passagens esquerda e direita, em função da posição silábica, por género.</i>	166
5.15	<i>Extensão das passagens laterais em (mm), valor médio para as passagens esquerda e direita, por género e vogal, para a posição de ataque.</i>	167
5.16	<i>Valores dos IP calculados para todos os pares de contornos, em função da posição silábica. Pi-Pm representa a comparação de contornos em posição de ataque (Pi) e ataque intervocálico (Pm), Pi-Pf representa a comparação de contornos entre a posição de ataque e de coda (Pf) e Pm-Pf a comparação dos contornos em posição de ataque intervocálico e coda.</i>	169
5.17	<i>Valores dos IP calculados para os pares de contornos [a-i], [a-u] e [i-u], para cada uma das posições silábicas.</i>	175
5.18	<i>Extensão dos contactos (em mm) na linha média referentes ao [L], para os 7 falantes.</i>	179
5.19	<i>Comprimento e área média, para as diferentes cavidades do [L], como produzido em [paL6] por quatro falantes; M (Masculino); F(Feminino).</i>	184
5.20	<i>Valores dos IP calculados, em posição intervocálica para o /L/, para os pares de contornos [a-i], [a-u] e [i-u]</i>	185
5.21	<i>Comparação dos valores das Formantes (F1, F2, F3) e zeros (em Hz), para o [L], obtidos nas simulações efetuadas com o VTAR e as suas produções naturais, para quatro falantes.</i>	188
5.22	<i>Comprimento médio do trato vocal, para seis dos falantes participantes no estudo, para as 9 produções, efetuadas para o /l/.</i>	191
5.23	<i>Comprimento médio do trato vocal para seis falantes participantes no estudo, em função da posição silábica.</i>	192
6.1	<i>Parâmetros de aquisição utilizados, com sequência TurboFLASH, com diferente resolução temporal.</i>	214
6.2	<i>Tabela resumo com os valores relativos à resolução temporal, resolução espacial e frames/s conseguidas para as várias séries.</i>	216
6.3	<i>Valores médios da duração (em ms) dos movimentos do véu palatino (abertura, fase estável, fecho e duração total), para as falantes CO e CM, em posição inicial (n = 10) e medial (n = 10).</i>	231

6.4	Valor médio, mediana, máximo, mínimo e desvio padrão para os valores de máxima abertura do PVF (mm^2), por falante.	241
6.5	Valor médio respeitante ao valor máximo do P_{NO} , por falante e para cada vogal nasal.	243
6.6	Valores médios (ms) e desvio padrão da distância temporal entre o target de abertura do véu palatino e o target do gesto oral (TTLag), para as posições inicial e medial.	245
6.7	Valores médios (ms) e desvio padrão da distância temporal entre o target de fecho do véu palatino e o gesto de fecho oral (T2TLag), para as posições inicial e medial.	245
F.1	Folha de registo das sessões de RM.	343

Capítulo 1

Introdução

The history of speech production studies can be understood to be the history of the development of observation techniques.

KIYOSHI HONDA (2002)

A fala, a forma natural de comunicação entre os seres humanos, representa a parte motora de um processo complexo que é a linguagem. No dia-a-dia, fazemos uso da linguagem e da fala sem dedicarmos muita atenção à forma como o fazemos. Por outras palavras, fazemo-lo sem necessidade de tomar consciência desse ato. No entanto, a subtileza dos movimentos dos articuladores¹ quando produzimos fala (e.g., movimentos da língua, lábios) estão entre os mais especializados do conjunto das ações efetuadas pelos seres humanos, dificultando o seu estudo e a sua modelação. A configuração dos diferentes articuladores e a forma do trato vocal assumida durante a produção dos diferentes sons é extremamente importante, mas representa apenas um dos diferentes níveis hierárquicos envolvidos na produção de fala, o que demonstra a sua grande complexidade² (Honda, 2003).

Devido à sua importância para a nossa vivência diária e pela complexidade que encerra, o estudo da fala e da sua produção tem merecido a atenção das mais variadas áreas do conhecimento, em que a área das ciências e tecnologias da saúde está incluída (e.g., imagem médica, anatomia e fisiologia, terapia fala), interligando-se, aos mais variados níveis, com as ciências exatas, com diferentes áreas da engenharia, da linguística e fonética, entre muitas outras (veja-se, por exemplo, o exposto em Fant, 2001).

Para além de estudar a fala, desde há muitos anos que é desejo do Homem conseguir criar e desenvolver máquinas que possam reproduzir a fala humana (síntese de fala). Apesar de todo

¹Os articuladores são as estruturas responsáveis pela articulação dos sons. Ao assumirem configurações diferentes, os articuladores produzem sons e regulam a produção de fala. Podem ser divididos em articuladores passivos ou ativos. Os ativos apresentam mobilidade e são a língua, os lábios, o palato mole e a mandíbula. Os articuladores passivos são os alvéolos dentários superiores, a porção anterior da arcada dentária superior e o palato duro (Andrade e Viana, 1996; Mateus *et al.*, 2005).

²Segundo Honda (2003, p.26) “...There are four hierarchical levels in speech production: the speech sound level, vocal tract shape level, vocal organ configuration level, and muscle contraction level. There is a “one-to-many” relationship among levels starting from the sound level and moving toward the muscle contraction level.”.

este interesse e da atenção que lhe tem sido conferida, ao longo dos tempos, utilizando diferentes abordagens, a síntese de fala não é, ainda, um desafio totalmente resolvido (Taylor, 2009), muito provavelmente porque ainda se tem conhecimento limitado acerca dos mecanismos inerentes à produção de fala. Dificilmente se conseguirá imitar algo que não se conhece bem:

“...I should have a perfect Knowledge of what I wanted to imitate (...). In this way my talking machine and my theory concerning speech made equal progress, the one serving as a guide to the other...” (afirmação de Von Kempelen 1791, referida em Teixeira, 2000)

O problema coloca-se, particularmente, para a síntese articulatória, uma das técnicas³ de síntese em que se procura simular um sinal de fala modelando as características físicas, anatómicas e fisiológicas dos órgãos que a produzem. Ou seja, modela-se o sistema em si e não o sinal produzido ou as suas características acústicas (Teixeira *et al.*, 2005; Birkholz *et al.*, 2007). Estes sintetizadores incluem, regra geral, dois subsistemas: um modelo anátomo-fisiológico das estruturas envolvidas na produção de fala e um modelo de produção e propagação do som nessas estruturas. O primeiro permite que se transforme a posição dos diversos articuladores durante a produção de sons em áreas transversais do trato vocal. O segundo modelo consiste num conjunto de equações que descrevem as propriedades acústicas do trato vocal (Teixeira, 2000). Outros parâmetros têm que ser adicionados para que se consiga modelar corretamente a produção de fala⁴. A nível mundial, não existem muitos sistemas⁵ que tenham por base a síntese articulatória. Na opinião de alguns autores (Engwall, 2002a; Teixeira, 2000; Taylor, 2009), um dos aspetos que mais tem condicionado a evolução dos sintetizadores de base articulatória reside na necessidade de obter bases de dados de produção de fala extensas e na própria dificuldade em obter esses dados. Um desses sintetizadores, o SAPWindows (Sintetizador Articulatório do Português para plataforma Windows) está em desenvolvimento, desde 1995, na Universidade de Aveiro (UA) por Teixeira (2000) e a sua evolução, tendo em vista o alargamento das suas capacidades, beneficiaria com a existência de dados de produção para o Português Europeu (PE).

1.1 Técnicas instrumentais para estudo e modelação da fala

Para obtenção destes dados de produção têm sido utilizadas variadas técnicas instrumentais, que detalharemos mais adiante, permitindo aumentar o conhecimento fonético e linguístico, para diversas línguas, e evoluir os modelos de síntese de fala. Como se refere na citação de abertura deste capítulo, a evolução dos estudos de produção de fala está intimamente relacionada com os desen-

³Na atualidade, coexistem várias técnicas de síntese de fala que incluem a síntese de formantes, a síntese por concatenação, a síntese articulatória, entre outras.

⁴Uma revisão detalhada acerca das questões envolvidas no desenvolvimento destes modelos pode ser encontrada em: Teixeira (2000); Teixeira *et al.* (2005); Taylor (2009).

⁵Exemplos de sintetizadores de base articulatória são: o sistema CASY (*Configurable Articulatory Synthesizer*) desenvolvido e evoluído nos laboratórios *Haskins*, nos EUA (Rubin *et al.*, 1981, 1996; Iskarous *et al.*, 2003); o ArtiSynth da *University of British Columbia* (Fels *et al.*, 2005; Van Den Doel *et al.*, 2006); o *Virtual Talking Head* (VTH), desenvolvido no *Institute de la Communication Parlé* em Grenoble (Badin *et al.*, 2000); o sintetizador articulatório 3D desenvolvido por Birkholz que, para além da voz falada, permite a síntese de voz cantada (Birkholz *et al.*, 2006; Birkholz, 2007, 2013).

volvimentos ao nível das técnicas de observação que, em algumas áreas, tem sido surpreendente (Honda, 2002).

Quando se pretende estudar a fala, é importante perceber que para que esta seja produzida contribuem diferentes articuladores que variam, não só na forma, estrutura e localização, mas também na dimensão e velocidade dos seus movimentos, dificultando a sua avaliação. A própria interação dos articuladores entre si, comparável a uma sofisticada coreografia, é difícil de analisar, mas extremamente relevante no mecanismo de produção e modelação de fala. Para Engwall (2002b), o ideal seria a existência de um único método que satisfizesse vários requisitos, nomeadamente, que incluísse todo o trato vocal, permitisse modelação tridimensional (3D), apresentasse elevada resolução temporal e espacial com boa relação Sinal-Ruído (doravante designada como S/R)⁶ e permitisse a aquisição simultânea de imagem e áudio. Para além disso deveria ser minimamente invasivo e permitir a obtenção de informação de forma segura.

No entanto, é difícil que uma técnica instrumental consiga, por si só, preencher todos os requisitos necessários para avaliar o intrincado processo que é a fala. Por esta razão, é frequentemente necessário recorrer a várias técnicas que, entre si, possam fornecer informação complementar (Stone, 1999; Engwall, 2002b) ou selecionar a que melhor se adequa à resolução de um particular problema ou questão. Ainda que este método ideal possa não existir, muitos autores consideram que, na atualidade, os avanços verificados a nível da Ressonância Magnética (RM) a posicionam como uma das mais promissoras na área (Bresch *et al.*, 2008a; Kim, 2010; Narayanan *et al.*, 2011).

Assim, nos parágrafos seguintes, providenciaremos uma visão muito global, que se complementará ao longo do restante trabalho, relativa à utilização de diferentes técnicas, referindo as suas potencialidades e constrangimentos, no âmbito da fonética articulatória.

Das várias técnicas instrumentais que são utilizadas para estudar a fala, poderemos referir como exemplos, a Articulografia Electromagnética 2D (EMA 2D)⁷ e, mais recentemente, EMA 3D, o sistema *X-Ray Microbeam* (XRMB) e a Electropalatografia (EPG). Mais dirigidas para o estudo dos sons nasais, refira-se também o sistema *Velotrace* (Bell-Berti *et al.*, 1993) e a Fibroscopia (Amelot *et al.*, 2003).

O sistema EMA 2D permite obter taxas de amostragem muito elevada (100-200 Hz), viabilizando a captação de características dinâmicas dos articuladores e extração de informação relativa à própria coordenação entre estes (coordenação intergestual). Esta técnica possibilita a aquisição de *corpora*⁸ extensos para vários falantes, num período de tempo relativamente curto, sendo muito útil, por exemplo, na avaliação de efeitos coarticulatórios ou variabilidade intra e inter sujeitos (Hoole e Nguyen, 1999; Mooshammer *et al.*, 2004; Recasens e Espinosa, 2009). No entanto, a informação obtida é condicionada a poucos pontos, na linha média, onde são colocados os sensores (geralmente três na língua, dois nos lábios, um no véu palatino⁹ e dois sensores de referência). A colocação de alguns sensores é extremamente difícil e, como o sensor mais posterior

⁶Do inglês *Signal-to-Noise Ratio SNR*.

⁷Neste trabalho utilizaremos a sigla EMA para designar esta técnica que se designa em língua inglesa por: *ElectroMagnetic Articulography*.

⁸Plural de *corpus* - conjunto de sons da fala, utilizado na investigação e desenvolvimento das aplicações em processamento de fala.

⁹Outras designações são véu do palato ou palato mole.

é colocado no véu palatino, não permite obter informação relativa à cavidade faríngea. Acresce ainda o facto de a colocação dos sensores, dentro da cavidade oral, poder perturbar a normal articulação dos sons. Os sistemas EMA 3D constituíram uma evolução significativa, permitindo obter informação para lá da linha média (Zierdt *et al.*, 1999; Hoole *et al.*, 2003), mas continua a ser um método em que a informação diz respeito a alguns pontos na cavidade oral, sendo ainda pouco acessível. Alguns estudos recentes utilizam esta técnica para estudo da nasalidade (Carignan *et al.*, 2011; Shosted, 2011; Shosted *et al.*, 2012a).

A EPG permite avaliar de forma precisa o local, a extensão e o *timing* exato dos contactos da língua com o palato, dentro da cavidade oral, sendo uma técnica relevante no estudo da articulação dos sons consonânticos (Recasens e Espinosa, 2009; Reis e Espesser, 2006) e, também, na avaliação de efeitos contextuais. Para o estudo dos sons vocálicos, a sua utilidade é, no entanto, questionável. Apresenta boa resolução temporal, todavia a necessidade de introduzir um palato artificial, que perturba a articulação, e o facto de não permitir estudar a cavidade faríngea, são apontadas como limitações ao método.

Algumas das técnicas acima referidas são invasivas (e.g., o *Velotrace*), outras são pouco acessíveis (e.g., XRMB) para a maioria dos investigadores, tornando-se dispendiosas, uma vez que são escassos os sistemas existentes a nível mundial. No caso do XRMB existe um equipamento na Universidade de Tóquio e outro, nos Estados Unidos, na Universidade de Wisconsin (Honda, 2002).

Para além destas, muitas modalidades de imagem médica têm dado um valioso contributo na área da produção de fala. Uma das mais paradigmáticas teorias na área - a teoria acústica de produção de Fala (Fant, 1960) - foi desenvolvida tendo por base imagens radiográficas de falantes do Russo, na sequência do trabalho pioneiro de Chiba e Kajiyama (1941) com vogais do Japonês¹⁰. Acompanhando sempre de perto as evoluções na área da imagem, também a Tomografia Computorizada (TC)¹¹, a Ultrasonografia (US) e a RM, mais conhecidas pelas suas potencialidades e utilizações na área do diagnóstico médico, têm sido “chamadas” para angariar conhecimento para a área da fonética, da modelação e síntese de fala.

No entanto, nos dias de hoje, conhecidos os efeitos biológicos da radiação ionizante, estando bem regulamentada e balizada a utilização de radiação ionizante - particularmente referida no artigo 3 da diretiva 43 da EURATOM (1997)¹² - não é aceitável, nem eticamente correta a sua utilização em indivíduos, sem qualquer tipo de perturbação, que não retirariam qualquer benefício de uma exposição, ainda que pequena, à radiação ionizante. Por esse motivo, o recurso às modalidades de imagem que utilizam o espectro ionizante (e.g., a cine-radiografia ou a TC) para estudar a fala dita “normal”, tem sido, gradualmente¹³, abandonado. Assim, neste domínio,

¹⁰Outros trabalhos que utilizaram imagens radiográficas obtidas com RX simples ou cine-radiografia são: Moll (1960); Stevens e Ohman (1963); Perkell (1971); Mermelstein (1973); Zerling (1984). Para o estudo dos sons do PE também Martins (1988) e Barbosa (1994) utilizaram traçados obtidos a partir de imagens radiográficas.

¹¹Uma das primeiras utilizações da TC nesta área foi efetuada por Sundberg *et al.* (1987) para obter áreas da faringe, em 4 vogais da língua Sueca. Tom *et al.* (1999) recorrem à tomografia por feixe de electrões, do inglês *Electron Beam Computed Tomography* (EBCT), num estudo de produção, na área do canto. Mais recentemente, Serrurier e Badin (2005b), num estudo de modelação, utilizam esta modalidade para extrair informação acerca das arcadas dentárias e palato duro de um falante.

¹²*European Atomic Energy Community* (EURATOM).

¹³Apesar das recomendações, infelizmente, ainda se continuam a utilizar estas técnicas de imagem, em voluntários,

tem-se assistido a uma crescente utilização das modalidades de imagem que não recorrem ao uso de radiações ionizantes, como é o caso da US e da RM.

A US tem merecido algum destaque nos últimos anos, principalmente, porque é uma técnica segura, barata, simples de executar, que permite efetuar gravação do sinal acústico sem necessidade de recorrer a tecnologia dispendiosa e apresenta uma resolução temporal adequada à captação do padrão de variação temporal dos articuladores, em tempo real (Stone, 2005; Wrench e Scobbie, 2011). Para além disso, a sua portabilidade favorece a utilização conjunta com outras técnicas como EPG ou Optopalatografia (OPG). Tem sido muito utilizada em estudos de coarticulação (Zarkova *et al.*, 2008), para efetuar reconstruções do dorso da língua (2D e 3D) para vários sons (e.g., Stone e Lundberg, 1996), caracterizar os gestos articulatorios das consoantes líquidas (e.g., Proctor, 2009), no contexto de *Silent Speech Interfaces-SSI* (Hueber *et al.*, 2008) e, também, como meio de obter *feedback* visual em crianças com alterações da fala (Modha *et al.*, 2008). No entanto, a sua utilização restringe-se ao estudo do dorso lingual (uma vez que o campo de visão nem sempre permite uma correta observação do ápice ou raiz da língua). Estruturas como o véu palatino e palato duro são mal visualizadas (Stone, 2005; Wrench e Scobbie, 2011). Também, para que não existam movimentos, é necessário efetuar imobilização da cabeça mediante a utilização de sistemas dedicados (e.g., sistema *HATS - Head and Transducer Support System*), nem sempre disponíveis, mas fundamentais nestes estudos (Stone, 2005). Uma revisão mais detalhada relativa à utilização destas e de outras técnicas exploratórias pode ser encontrada em Stone (1999), Engwall (2002b) e Martins (2007).

A RM é uma modalidade de imagem que apresenta capacidade multiplanar, permite a obtenção de informação relativa ao trato vocal em toda a sua extensão, incluindo a cavidade faríngea (inacessível a outras técnicas), de forma segura e não invasiva, permite modelação 2D e 3D, apresenta uma resolução de contraste a nível dos tecidos moles elevada, permitindo uma excelente visualização de articuladores como a língua, o véu palatino e os lábios. Algumas das limitações¹⁴ associadas à RM, no âmbito dos primeiros estudos de produção da fala, têm vindo a ser minimizadas ou mesmo ultrapassadas¹⁵ fazendo com que esta modalidade de imagem se profile como uma das mais promissoras nesta área e da qual muito ainda se pode esperar (Bresch *et al.*, 2008a; Kim, 2010; Narayanan *et al.*, 2011).

Na realidade, os desenvolvimentos tecnológicos observados na área da RM - a que conferiremos algum destaque no capítulo 2 - permitem, na atualidade: 1) a obtenção de volumes de dados, com elevada resolução espacial, em escassos segundos; 2) a aquisição de imagens em tempo real com uma taxa de amostragem que permite, de forma satisfatória, captar a dinâmica do trato

para estudar a fala.

¹⁴Como limitações da RM, no âmbito da produção de fala, são geralmente apontadas: i) a baixa resolução temporal, quando comparada com outros métodos, ii) o ruído produzido pelas bobinas de gradientes impedindo ou dificultando a gravação do sinal acústico durante as aquisições de imagem, iii) o facto de a aquisição ser efetuada com o falante em supinação, tendo que ter em conta, na interpretação dos resultados, alguns efeitos gravitacionais (Tiede *et al.*, 2000; Engwall, 2000b), iv) a necessidade de, nos estudos estáticos, o falante ter que sustentar as articulações durante o período aquisição, conduzindo a produções hiperarticuladas (Engwall, 2000b) e, v) má visualização das estruturas calcificadas, dificultando a observação dos dentes e do palato (Takemoto *et al.*, 2004; Narayanan *et al.*, 1997b).

¹⁵A utilização, cada vez mais generalizada, de microfones óticos (NessAiver *et al.*, 2006; Bresch *et al.*, 2006b; Niebergall *et al.*, 2012) permitindo a aquisição sincronizada de som e imagem durante as aquisições de RM, permite superar uma das limitações mais apontadas, constituindo uma mais-valia nestes estudos.

vocal e observar a interação entre os diversos articuladores durante a produção de fala (Demolin *et al.*, 2000; Narayanan *et al.*, 2004; Bresch *et al.*, 2008a); 3) com compromissos a nível da resolução temporal e com uma qualidade de imagem ainda baixa é já viável a aquisição de volumes de dados (RM-3D), em tempo real (Zhu *et al.*, 2013b).

Desta forma, tem sido possível alargar o espetro de sons que se podem estudar com RM, como é o caso de sons com características mais dinâmicas como os sons nasais (Byrd *et al.*, 2009; Kim *et al.*, 2010) ou os róticos (Nishimoto *et al.*, 2011); adquirir imagens durante a produção de palavras reais ou mesmo frases (Narayanan *et al.*, 2011), ao invés de sons artificialmente sustentados; contemplar aspetos prosódicos e emoções (Lee e Narayanan, 2010; Ramanarayanan *et al.*, 2010); adquirir *corpora* mais extensos (Narayanan *et al.*, 2011) e, ainda, estudar aspetos de coordenação intergestual (Byrd *et al.*, 2009; Lammert *et al.*, 2010a; Proctor *et al.*, 2013).

Pelo valor que lhe tem sido reconhecido nesta área (fruto da sua versatilidade) a RM, através do recurso a diferentes técnicas (estática, dinâmica, em tempo real, *tagged*) tem sido amplamente utilizada em estudos de produção de fala, para várias línguas, com diferentes propósitos (veja-se alguns exemplos no capítulo 2). Para o Português Europeu, a utilização de RM no contexto de estudos de produção de fala é escassa, relativamente recente, tendo sido dados os primeiros passos com os estudos realizados por Martins (2007) e Rua e Freitas (2007).

1.2 Problemas e motivações

Apesar de ter existido algum investimento recente a nível da recolha de dados articulatórios para o PE e também algumas evoluções importantes no SAPWindows, conseguidas na sua maioria no âmbito do projeto HERON¹⁶, a verdade é que continua a existir escassez ou mesmo inexistência de dados de produção, particularmente, para algumas classes de sons desta língua. A afirmação de Oliveira (2009), que a seguir se transcreve, ilustra bem e fundamenta a necessidade de obter mais dados de produção:

“... A recolha de dados de produção é absolutamente essencial para a validação e evolução das propostas de caracterização gestual dos vários segmentos do PE. O problema é particularmente premente em relação a algumas classes de sons, como as laterais alveolares ou as vibrantes, cujas configurações gestuais carecem reconhecidamente de ser aferidas e, se necessário revistas.” (Oliveira, 2009, p.245)

¹⁶O projeto HERON financiado pela FCT (POSC-PLP/57680/2004) - *A Framework for Portuguese Synthesis Research* iniciou-se em 2004 com objetivo de: a) adquirir dados de produção da fala, b) permitir a evolução do SAPWindows para inclusão de outros sons (e.g., fricativas) e c) desenvolvimento dos módulos linguísticos com o intuito de obter um sistema completo (Texto-Fala) de base articulatória. Na fase inicial deste projeto, estava prevista a aquisição de dados de RM na Universidade da Florida, nos Estados Unidos. No que diz respeito à aquisição de dados, em 2006, reuniram-se as condições para aquisição dos dados de RM, em Portugal, no âmbito do trabalho de Mestrado da autora da presente investigação. A aquisição de imagens foi conduzida nos Hospitais da Universidade de Coimbra. O estudo incluiu um falante com um *corpus* muito genérico incluindo a maior parte dos sons do PE. Em simultâneo, foram desenvolvidas algumas rotinas de segmentação em imagens 2D e 3D, que viabilizaram a extração de informação articulatória (Martins, 2007; Martins *et al.*, 2008; Carbone, 2008). Posteriormente, procedeu-se à aquisição de dados RM, para mais dois falantes, e recolha de dados EMA, em *Grenoble*, com intuito de viabilizar a primeira proposta de um modelo gestual para todos os sons do PE, que viria a ser apresentada por Oliveira (2009). No seu conjunto, estes dados têm permitido a evolução de diferentes módulos do SAPWindows.

De facto, no que diz respeito às consoantes laterais, para além da falta de dados articulatórios, há ainda muitos aspetos teóricos relativos à sua produção, particularmente, da lateral alveolar /l/, que são alvo de alguma controvérsia.

Para os sons róticos¹⁷ não existem, que seja do nosso conhecimento, dados articulatórios e os acústicos são escassos (Jesus e Shadle, 2005).

Uma outra classe de sons com grande relevo, em particular, para o Português, são as vogais nasais. As suas características específicas fazem com que exista a necessidade de obtenção de mais informação relativa à sua produção. Não estão ainda clarificadas muitas questões importantes nomeadamente, aspetos como a variação do véu palatino no tempo, a coordenação entre os gestos¹⁸, e outros aspetos dinâmicos que têm a ver, por exemplo, com a existência de diferentes fases ao longo da sua produção. Ao contrário das laterais e dos róticos, existem, para as vogais nasais do PE, variados estudos (Teixeira, 2000; Teixeira e Vaz, 2001; Rossato *et al.*, 2006; Oliveira, 2009) que abordam alguns dos problemas enumerados.

No entanto, há limitações na informação atualmente disponível, como o número restrito de falantes, *corpus* RM não específicos para algumas classes de sons, informação demasiado estática e, maioritariamente, 2D ou informação dinâmica limitada a pontos ao longo do trato vocal.

Desta forma, este trabalho surge de forma natural, como sequência do trabalho encetado no âmbito do Mestrado em Ciências da Fala e Audição onde se recorreu à RM para obtenção de dados de produção para o PE e enquadrando-se dentro dos objetivos, mais abrangentes, do projeto HERON II¹⁹ visando colmatar esta escassez de dados tão referida para o PE, em particular, para as consoantes laterais, angariar informação complementar relativa às vogais nasais e contribuir para um melhor conhecimento da produção destes sons.

Dar continuidade a uma linha de investigação, já bem consolidada, no seio de uma equipa multidisciplinar e poder evoluir o trabalho iniciado anteriormente constituíram as nossas principais motivações. Poder fazê-lo recorrendo a uma modalidade imagiológica (a RM) com a qual já trabalhamos enquanto profissionais de Radiologia, é ainda mais aliciante.

Para além dos desafios colocados nas tarefas de otimização e aquisição de imagem, o envolvimento direto da autora nas tarefas de processamento de imagem, que não se verificou durante o trabalho de Mestrado, é também fator de motivação uma vez que será necessário desenvolver algumas competências nessa área. O processamento de imagem é um passo extremamente relevante

¹⁷Os sons róticos como o [ʀ] e [ʁ/] são consoantes produzidas com uma obstrução parcial no trato vocal que provoca a vibração de um articulador. Na terminologia portuguesa, são usualmente designados consoantes líquidas vibrantes (Mateus *et al.*, 2005).

¹⁸No Inglês corresponde a *gesture*. No domínio da Fonologia Articulatória a expressão gesto tem um significado próprio: “In Articulatory Phonology, the basic phonological units are gestures...Each gesture corresponds to the formation of a constriction within one of the relatively independent articulatory subsystems of the vocal tract and can be modeled as a task-dynamic system or control regime.” (Browman e Goldstein, 1995 p.19).

¹⁹O projeto Heron II - *Evolution of the European Portuguese Articulatory based Text-to-Speech Framework*, Ref. PTDC/EEA-PLP/098298/2008 - dá continuação ao Projeto HERON (Teixeira *et al.*, 2009), concluído em 2007 em que também participamos. O HERON II apresenta 3 grandes linhas de investigação: 1) aquisição de novos dados de produção (principalmente obtidos através de imagem em tempo real), que suportem o desenvolvimento de novos modelos articulatórios para o PE; 2) evolução progressiva do sintetizador de base articulatória da Universidade de Aveiro (SAPWindows) - inclusão de módulo para síntese das laterais; 3) aplicação de novas teorias linguísticas e da informação proveniente da análise dos dados de produção ao desenvolvimento de sistema completo de síntese (TTS), de base articulatória.

para a obtenção de dados. No entanto, a quantidade de informação, a complexidade geométrica do aparelho fonador e a própria qualidade das imagens, relativamente baixa, devido aos tempos de aquisição curtos, condicionam as opções e dificultam a sua implementação.

Não sendo frequente, no nosso país, o recurso às modalidades de imagem no contexto da produção e modelação de fala, dar a conhecer, junto dos profissionais e investigadores desta área, o potencial de aplicação das modalidades imagiológicas, em particular da RM, neste domínio, constitui-se como uma motivação adicional.

Apesar deste estudo não versar de forma direta o estudo das alterações da fala ou das patologias relacionadas com o trato vocal, poderá vir a ter implicações a esse nível. É extremamente importante conhecer a fala “normal” para auxiliar, na prática clínica, o diagnóstico das alterações e o planeamento de intervenções. Pensamos que o trabalho que aqui se descreve poderá ter aplicação em contexto clínico, uma vez que o objeto de estudo (e.g., aparelho fonador) e as metodologias usualmente utilizadas para a sua avaliação (e.g., modalidades de imagem) são comuns. A avaliação de patologias do foro articulatório (e.g., dislalia, disartria), alterações que podem interferir com a produção da fala (e.g., insuficiência/incompetência velofaríngea, macroglossia) e patologias relacionados com deglutição/disfagia ou apneia do sono podem vir a beneficiar dos protocolos otimizados no contexto do presente trabalho, podendo motivar e viabilizar a realização de trabalhos com profissionais de saúde de outras áreas como, por exemplo, terapia da fala ou otorrinolaringologia.

Tendo em atenção o exposto, fundamentados pela opinião de outros autores e pela nossa experiência prévia nesta área, defendemos em tese, que a utilização de RM, mediante o recurso a diferentes técnicas de aquisição de imagem (2D, 3D e tempo-real), com protocolos otimizados, coadjuvada com a exploração de técnicas de processamento adequadas às características das diferentes imagens, se posiciona, ainda que com algumas limitações, como uma estratégia adequada para a obtenção de dados articulatorios relevantes, atualmente inexistentes para o PE.

No seu conjunto, a abordagem proposta permitirá colmatar a falta de dados, para algumas classes de sons, e complementar informação já existente, visando uma contribuição para o esclarecimento de algumas questões relativas à produção dos sons do PE, nomeadamente, das consoantes laterais e vogais nasais. Para além dos contributos para uma melhor conhecimento fonético que se almejam alcançar, no âmbito deste trabalho, a informação recolhida terá relevância para o processo de evolução dos modelos que sustentarão o aperfeiçoamento de um sintetizador completo de síntese, de base articulatória, em desenvolvimento na UA por Teixeira (2000).

1.3 Objetivos

O nosso objetivo geral é a criação de uma plataforma de aquisição e processamento de imagens de RM, tendo em vista dar resposta a questões mais prementes, derivadas da escassez de dados articulatorios para o Português (e.g., laterais e vogais nasais), mas que possa ser facilmente estendida e contemplar, no futuro, o estudo de outros sons ou questões teóricas, a modelação 3D dos vários articuladores envolvidos na produção fala e, a mais longo prazo, a avaliação de perturbações

diretamente relacionadas com a produção de fala ou com o trato vocal. Para atingir este objetivo mais abrangente, definiram-se objetivos concretos que passamos a explicitar:

1. **Implementar protocolos de RM ajustados à produção de fala** - Antes de efetuar a aquisição de imagens é necessário criar as condições técnicas adequadas, contornando e resolvendo alguns constrangimentos, em função das diferentes questões que se pretendem estudar. Uma parte importante deste trabalho, talvez a menos visível, mas de maior importância, passa por, numa primeira fase, proceder à otimização de sequências, rotinas (e.g., sincronização som-imagem com microfones dedicados) e procedimentos, para obter protocolos de aquisição de imagem que possam ser utilizados no domínio da produção de fala. Pretende-se que estes protocolos possam vir a ser utilizados, de forma mais abrangente, num futuro próximo.

2. **Adquirir bases dados de produção** - A obtenção de dados articulatórios será fundamental para colmatar a escassez de dados existente para o PE e para permitir a evolução dos modelos de produção atualmente existentes. Bases de dados de produção são mais que um conjunto de imagens do trato vocal. Para que se adquira informação válida, no âmbito de produção de fala, torna-se necessário elaborar *corpora* ajustados e dimensionados, em função da técnica de aquisição que se utiliza, e considerando as questões específicas que se pretendem abordar. A elaboração ou ajuste de *corpora* já existentes constitui também um objetivo do trabalho.

Para estudar, descrever e modelar as consoantes laterais, é fundamental, neste momento, obter informação relativa à variação das áreas ao longo do trato vocal, à dimensão das passagens laterais, à forma da língua, às áreas de contacto, etc. Ou seja, para que se consiga efetuar uma descrição articulatória destas consoantes e sua posterior modelação é necessário obter **informação tridimensional**. Para atingir este objetivo a utilização de aquisições volumétricas (**RM-3D**) será, na nossa opinião, a melhor abordagem. A obtenção de imagens das arcadas dentárias dos falantes e co-registo com os volumes obtidos em fonação, fundamental para o estudo das consoantes laterais, deverá ser conseguida sem recurso à utilização de radiação ionizante.

A aquisição de **informação temporal**, por outro lado, é outra das metas a atingir, neste caso diretamente aplicável ao estudo das vogais nasais, mas, no futuro, extensível a *corpora* mais alargados de sons róticos e laterais. A RM obtida em **tempo real**, com uma resolução temporal adequada, deverá permitir extrair essa informação, com visualização de todo o trato vocal e não condicionada a um conjunto de pontos, permitindo ultrapassar alguns dos problemas existente com os dados atualmente disponíveis para o PE (e.g., obtidos com EMA). Para além disso, pretende-se a aquisição de dados novos de produção, só possíveis de obter através de RM.

3. **Contribuir para a implementação de estratégias de processamento de imagem adequadas às necessidades decorrentes do estudo da produção de fala** - É nosso intuito desenvolver competências neste domínio, explorando várias possibilidades de processamento para os diferentes tipos de imagem ou estruturas, recorrendo, sempre que possível,

a abordagens mais automatizadas, uma vez que a quantidade de dados a processar é muito elevada (neste estudo serão milhares de imagens).

4. **Obter informação para a área da fonética** - Pretendemos demonstrar de que forma os dados de produção obtidos com RM poderão acrescentar algum conhecimento fonético relativo à produção das laterais e das vogais nasais. Por exemplo, no que concerne às laterais, pretende-se efetuar uma descrição articulatória detalhada, que suportará também a construção de um modelo 3D, que permita vir a sintetizar, com qualidade, estes sons com o SAPWindows.

1.4 Resultados publicados

Ao longo do período em que decorreu o presente trabalho de investigação, foram apresentados e publicados alguns artigos que espelham também os contributos do presente trabalho em diferentes áreas. Enquadram-se, maioritariamente, em três grandes vertentes: 1) Aquisição de dados, 2) Processamento de imagem 3) Contributos fonéticos/modelação.

1. Aquisição de dados

- A descrição do método de aquisição, dos procedimentos para aquisição de imagem volumétrica e das técnicas de processamento utilizadas, para obter informação relativa às consoantes laterais, é apresentada em:

Martins, P., Oliveira, C., Ferreira, C., Silva, A., Teixeira, A. (2011). *3D MRI and semi-automatic segmentation techniques applied to the study of European Portuguese laterals*. In Proceedings of the 9th International Speech Production Seminar (ISSP 2011), p. 235-242, Montreal, Canadá.

- De forma semelhante, a descrição dos procedimentos seguidos na obtenção das imagens RM em tempo real, do *corpus*, dos primeiros resultados obtidos (para 1 falante) e, possíveis aplicações do método, está incluída num capítulo de livro:

Teixeira, A., **Martins, P.**, Oliveira, C., Ferreira, C., Silva, A., Shosted, R. (2012). *Real time MRI for Portuguese: Database, Methods and Applications*. Computational Processing of the Portuguese Language, vol. LNCS (7243/2012), p. 306-317, Springer.

2. Processamento de Imagem

- Os resultados do estudo de avaliação preliminar relativa à segmentação da língua e o recurso a ferramentas de computação gráfica aplicadas ao estudo dos sons do PE deram origem a duas publicações em conferência:

Martins, P., Oliveira, C., Silva, S., Silva, A., Teixeira, A. (2011). *Tongue segmentation from MRI images using ITK-SNAP: Preliminary Evaluation*. In proceedings of the IADIS CGVCVIP-2011, p. 3-10, Roma, Itália.

Martins, P., Silva, S., Oliveira, C., Ferreira, C., Silva, A., Teixeira, A. (2011). *Investigating the differences between European Portuguese sounds: An Approach using Polygonal Mesh Comparison*. In F. Silva, D. Gutierrez, J. Rodríguez, M. Figueiredo (Eds). V Ibero-American Symposium in Computer Graphics SIACG-2011, p. 209-216, Faro, Portugal.

Na sequência da apresentação efetuada no SIACG surgiu o convite, por parte dos editores, para estender o trabalho apresentado na conferência, para um artigo em revista. Nesta publicação, demonstra-se que o recurso a uma ferramenta de comparação de malhas, usualmente utilizada para outros fins, pode trazer alguns contributos para a área da fonética. Constitui uma abordagem inovadora com várias possibilidades de aplicação (e.g., para comparar diferentes sons, o mesmo som produzido em contextos ou posições diferentes, para avaliar diferenças entre falantes, entre outras).

- **Martins, P.**, Oliveira, C., Silva, S., Ferreira, C., Silva, A., Teixeira, A. (2012). *Polygonal Mesh Comparison applied to the study of European Portuguese Sounds*. International Journal of Creative Interfaces and Computer Graphics (3), p. 28-44, IGI-Global.
- O método de segmentação aplicado aos cortes coronais, com o intuito de obter a variação das áreas orais e nasais, ao longo do tempo, é apresentado em:

Silva, S. **Martins, P.**, Oliveira, C., Silva, A., Teixeira, A. (2012). *Segmentation and Analysis of the Oral and Nasal Cavities from MR Time Sequences*. In proceedings of 9th International Conference on Image Analysis and Recognition (ICIAR 2012), Aveiro, Portugal.

3. Contributo para conhecimento fonético

Nesta secção, incluem-se todas as publicações em que são apresentados resultados com interesse para o estudo e modelação das consoantes laterais, das vogais nasais e incluem-se, também, as publicações relacionadas com um estudo de coarticulação.

- Os primeiros resultados obtidos com o estudo estático, relacionado com as consoantes laterais, foram apresentados na conferência FALA 2010. Nesta publicação, é apresentada informação 2D relativa aos sete falantes participantes no estudo e informação 3D para dois dos falantes. Este trabalho foi contemplado com um *Best Student Paper Award*.

Martins, P., Oliveira, C., Silva, A., Teixeira, A. (2010). *Articulatory Characteristics of European Portuguese laterals: a 2D and 3D MRI study*, In C. G. Mateo, F. C. Díaz and F. M. Pazó (eds.). FALA 2010: VI Jornadas en Tecnologia del Habla, Speech and Language Technologies for Iberian Languages, Vigo, Spain.

- A descrição articulatória e acústica da lateral alveolar /l/ deu origem a uma publicação no *International Conference of Phonetic Sciences* (ICPhS-2011). Nesta fase do nosso trabalho, já grande parte dos dados 3D tinham sido alvo de análise. Os dados acústicos, incluídos na publicação, não foram recolhidos no âmbito do presente trabalho.

Oliveira, O., **Martins, P.**, Marques, I., Couto, P., Teixeira, A. (2011). *An articulatory and acoustic study of the European Portuguese /l/*. In proceedings of the 17th International Conference of Phonetic Sciences (ICPhS 2011), Hong Kong, China.

- Num capítulo de livro, são apresentados os resultados de uma experiência de modelação acústico-articulatória efetuada para a lateral palatal [L]. Para além dos resultados da modelação, inclui a descrição articulatória desta consoante.

Teixeira, A., **Martins, P.**, Oliveira, C., Silva, A. (2012). *Production and Modeling of the European Portuguese Palatal Lateral*. Computational Processing of the Portuguese Language, vol. LNCS (7243/2012), p. 318-328, Springer.

- O estudo dinâmico relativo à configuração oral das vogais nasais do PE, com informação relativa aos três falantes adquiridos, foi apresentado numa sessão especial da conferência *Interspeech 2012* intitulada “Vowel nasalization”. Os dados aí apresentados resultam das aquisições em tempo-real e permitem complementar, com informação dinâmica, os dados obtidos num estudo prévio com RM estática (Oliveira, 2009).

Oliveira, C., **Martins, P.**, Silva, S., Teixeira, A. (2012). *An MRI study of the oral articulation of European Portuguese nasal vowels*. In proceedings of 13th Annual Conference of the International Speech Communication Association (Interspeech 2012), Portland, USA.

- A exploração dos cortes coronais oblíquos, efetuados na zona velar, com o intuito de estudar os movimentos deste articulador, permitiram-nos quantificar aspetos temporais relacionados com o movimento de abertura e fecho do véu palatino, complementando os dados EMA já existentes para o PE e também, obter novos dados de produção como a variação das áreas orais e nasais ao longo do tempo e a proporção destas áreas na zona de acoplamento das cavidades oral e nasal.

Martins, P., Oliveira, C., Silva, S., Teixeira, A. (2012). *Velar movements in European Portuguese nasal vowels*. In proceedings of IberSpeech 2012, Madrid, Espanha.

- Finalmente, duas publicações em que são apresentados resultados relativos à avaliação de efeitos coarticulatórios em várias consoantes do PE.

Martins, P., Carbone, I., Silva, A., Teixeira, A. (2011). *Effects coarticulatoires sur le portugais européen: une première étude IRM*. In Embarki & Dodane (eds): *Production et Perception de la Parole: La coarticulation, des Indices aux Représentations*, pages 101-115, L’Harmattan.

Teixeira, A., **Martins, P.**, Oliveira, C., Silva, A. (2011). *An MRI study of consonantal coarticulation Resistance in Portuguese*. In Proceedings of the 9th International Speech Production Seminar (ISSP 2011), pages 243-250, Montreal, Canadá.

Outros artigos relacionados com o trabalho de Doutoramento

- As publicações seguintes estão relacionados com o trabalho de Doutoramento por razões diferentes: os dois primeiros dizem respeito ao envolvimento da autora na recolha de dados articulatórios com EMA para as consoantes laterais e para as vogais nasais, no contexto da sua participação no projeto HERON. Os outros dois artigos resultaram da análise dos dados

recolhidos no âmbito do trabalho de Doutoramento, constituindo evoluções do trabalho agora apresentado, na área do processamento e visualização dos dados.

Oliveira, C., Teixeira, A., **Martins, P.** (2009). *Towards an articulatory characterization of European Portuguese lateral /l/*. In proceedings of the ISCA tutorial and research Workshop on Experimental linguistics (ExLing 2010), p. 133-136, Atenas, Greece.

Oliveira, C., **Martins, P.**, Teixeira, A. (2009). *Speech rate effects on European Portuguese Nasal Vowels*. In proceedings of the 10th International Conference of the Speech Communication Association (InterSpeech 2009), Brighton, United Kingdom.

Samuel Silva, António Teixeira, Catarina Oliveira, **Paula Martins** (2013). *Segmentation and Analysis of Vocal Tract from MidSagittal Real-Time MRI*, In Proceedings of the 9th International Conference on Image Analysis and Recognition (ICIAR 2013), LNCS 7950, p. 459–466.

Samuel Silva, António Teixeira, Catarina Oliveira, **Paula Martins** (2013), *Towards a Systematic and Quantitative Analysis of Vocal Tract Data*, In Proceedings of InterSpeech 2013, p. 1307–1311.

1.5 Estrutura da tese

Na sequência deste capítulo introdutório, que pretende enquadrar esta investigação e dar conta das principais motivações e objetivos subjacentes à sua elaboração, este documento encontra-se estruturado em mais seis capítulos.

Assim, no **capítulo 2**, apresenta-se o estado da arte, no que diz respeito à modalidade de imagem utilizada, a RM, e sua aplicação específica na área da produção de fala. Proporciona-se uma visão global dos avanços tecnológicos da RM que têm vindo a viabilizar a implementação de protocolos, cada vez mais ajustados aos requisitos exigidos no âmbito da produção de fala. Uma vez que, para que os nossos objetivos sejam atingidos, é necessário não só a obtenção de imagens, como também o seu processamento apresenta-se, no **capítulo 3**, uma síntese das diferentes abordagens, atualmente disponíveis no campo da segmentação de imagem. De seguida, faz-se uma resenha das técnicas que mais têm sido utilizadas para processamento de imagens de RM, na área da produção de fala. No final, elencam-se os problemas e dificuldades neste domínio e apresentam-se as opções tomadas, neste estudo, para a exploração das bases de dados recolhidas.

Seguidamente, no **capítulo 4**, porque os sons que nos propomos estudar, no contexto deste trabalho, são as consoantes laterais e as vogais nasais, efetuar-se-á um enquadramento genérico relativo às principais características articulatórias destes sons, referindo-se o conhecimento atual no que diz respeito ao PE.

No **capítulo 5**, concentramo-nos na apresentação de todo o trabalho desenvolvido, com o intuito de contribuir para a caracterização articulatória das consoantes laterais e sua futura modelação. Depois de elencar as questões que serão analisadas, explicitam-se os objetivos a atingir com esta parte do nosso estudo. Segue-se uma descrição pormenorizada dos métodos e procedi-

mentos utilizados para obtenção da informação, onde se inclui a definição do *corpus*, dos falantes, dos procedimentos e das opções efetuadas relativamente à aquisição e processamento de imagem. Reportam-se, também, as tarefas de avaliação efetuadas e as experiências conduzidas com algumas ferramentas de visualização. Na secção seguinte do mesmo capítulo, são apresentados os resultados obtidos. A informação recolhida permitir-nos-á apresentar uma descrição articulatória detalhada da lateral alveolar e da palatal, que contempla aspetos como ponto de articulação, extensão do contacto, forma geral da língua, áreas e funções de área do trato vocal, passagens laterais e avaliação de efeitos posicionais/vocálicos na produção das laterais. Na parte final do capítulo, faz-se a discussão dos resultados, tendo em atenção os objetivos propostos e enquadrando-os com o conhecimento atual para o PE e para outras línguas.

O **capítulo 6**, centra-se nas aquisições de RM em tempo real e segue uma estrutura semelhante à do capítulo precedente, neste caso, direcionado para os aspetos relativos à produção das vogais nasais do PE. No início do capítulo, é feita uma contextualização do tema através da apresentação de estudos similares realizados para outras línguas (e.g., Francês e Português Brasileiro), identificam-se os problemas ainda em aberto e os objetivos que nos propomos alcançar com esta parte da nossa investigação. Descrevem-se, de seguida, as metodologias utilizadas onde se inclui a descrição do *corpus*, os falantes, a parte técnica relativa à recolha de dados e das abordagens de processamento utilizadas para extrair informação das imagens obtidas em tempo real. O capítulo culmina com a apresentação e discussão dos resultados que, mais do que significado fonético, pretendem demonstrar que tipo de informação e de que forma pode ser obtida a partir da base de dados recolhida.

Finalmente, no **capítulo 7**, são sintetizados os principais contributos deste estudo, são apresentadas as conclusões deste trabalho, considerando as suas implicações para a área da fonética/linguística e sua futura aplicação para o processo de síntese articulatória do PE. São apontadas algumas sugestões para desenvolvimentos futuro que poderão complementar e evoluir o trabalho agora apresentado.

Capítulo 2

Ressonância Magnética

Nothing is “free” in MRI.

STAFFORD (2004)

Neste capítulo, temos por objetivo efetuar um enquadramento teórico relativo à modalidade imagiológica que será utilizada no presente estudo: a Ressonância Magnética¹. Focalizaremos a nossa atenção nos principais avanços tecnológicos que se têm observado na área, contribuindo para que a RM se afirme, no presente, como uma das mais relevantes modalidades de imagem médica, permitindo obter, para além de informação morfológica detalhada, informação funcional relevante para a caracterização das mais diversas patologias. Estes avanços, particularmente a drástica diminuição nos tempos de aquisição, tornando possível, por exemplo, a obtenção de imagem em tempo real e aquisição rápida de volumes de alta resolução, têm permitido que a RM se torne uma ferramenta fundamental, cada vez mais utilizada na área da produção de fala.

Assim, providenciar-se-á uma visão global desta evolução, referindo aspetos relacionados com a intensidade do campo magnético, com os gradientes de campo, com as antenas de Radiofrequência (RF) e com outras técnicas, que têm vindo a ser implementadas, que visam acelerar o processo de aquisição de imagem e melhorar a sua qualidade.

Finalmente, porque neste trabalho a RM é utilizada no contexto da produção de fala, elencar-se-ão alguns estudos, dos muitos já realizados neste domínio, ilustrando de que forma os progressos observados no seio desta modalidade imagiológica têm sido capitalizados para a área da produção de fala.

A RM é uma modalidade de imagem tomográfica que, à semelhança de outras, consegue informação que representa a distribuição espacial de uma grandeza física. Apresenta capacidade multiplanar, ou seja, permite obter informação bidimensional (2D) de forma direta, em qualquer orientação, informação volumétrica (3D) e informação (4D) no caso de volumetria dinâmica ou

¹Uma revisão pormenorizada relativa aos princípios físicos básicos desta técnica não se enquadraria no âmbito deste trabalho, pelo que remetemos o leitor, mais interessado nesta matéria, para algumas referências consideradas pertinentes nesta área do conhecimento como, por exemplo: Liang e Lauterbur (2000), Elster e Burdette (2001), Brown e Semelka (2010) ou Hashemi *et al.* (2012).

se for considerada uma distribuição espectro-espacial (Liang e Lauterbur, 2000). Para além disso, permite explorar vários mecanismos de contraste em função de parâmetros intrínsecos dos próprios tecidos (e.g., densidade protónica) e de fatores extrínsecos (parâmetros de aquisição). Proporciona, por isso, uma riqueza de informação dificilmente conseguida com outras modalidades de imagem. A elevada resolução de contraste, particularmente a nível dos tecidos moles, é uma das suas mais apontadas vantagens. É considerado um método de aquisição de imagem seguro², uma vez que utiliza o espectro eletromagnético na gama da RF, tornando-o, por essa razão, muito apelativo como ferramenta de investigação no estudo e modelação da fala, uma vez que a maior parte dos estudos, nesta área, são conduzidos em voluntários sem qualquer tipo de patologia.

Para a obtenção de imagens por RM é necessário: (1) a presença de um campo magnético estático, homogêneo e uniforme, com uma intensidade B_0 produzido por um magnete, (2) um sistema de RF para emissão/receção de sinal (antenas de RF, também designadas por bobines ou *coils*) e (3) bobines de gradientes que originam um campo magnético variável (dB/dt), que se sobrepõe a B_0 e que permite efetuar a codificação do sinal. É necessário ainda um sistema computadorizado e *hardware* e *software* adequados para aquisição e processamento da informação (Liang e Lauterbur, 2000; Elster e Burdette, 2001; Huettel *et al.*, 2004).

Desde a sua introdução na prática clínica, nos finais da década de 70, altura em que os tempos de aquisição eram muito longos e as suas aplicações mais restritas, muitos têm sido os desenvolvimentos tecnológicos aos mais variados níveis, que permitem, no presente, a captura de imagens em tempo real com elevada resolução temporal (e.g., Uecker *et al.*, 2010b; Zhang *et al.*, 2012 e Zhu *et al.*, 2013a) e obtenção de volumes de dados com resolução isotrópica em poucos minutos (Gold *et al.*, 2007; Edelman *et al.*, 2009). Com algum sacrifício na resolução espacial, é viável obter volumes de dados consideráveis com resolução quase isotrópica em escassos segundos (e.g., Martins *et al.*, 2008 ou Kim *et al.*, 2009)³.

No momento presente, apesar de existirem, não seriam os constrangimentos de ordem técnica e tecnológica que inviabilizariam ir mais além no que diz respeito, por exemplo, à maior rapidez de aquisição. No entanto, como se evidenciará, começam a ser atingidos alguns limiares fisiológicos levando a que outras estratégias tivessem que emergir, por forma a viabilizar aquisições mais rápidas, sem grande sacrifício da qualidade de imagem, de uma forma segura (Tsao *et al.*, 2003; Stafford, 2004; Lustig *et al.*, 2008).

²Consideramos pertinente afirmar que para que um exame de RM seja efetuado de forma segura têm que ser observadas várias normas e procedimentos de segurança, sem as quais poderão ocorrer situações indesejáveis (Shellock, 2000, 2001) devendo ser seguidas as diretivas emanadas por várias entidades reguladoras como a *Food and Drug Administration (FDA)*, *Medicine and Healthcare products Regulatory Agency (MHRA)*, *Health Protection Agency (HPA)*, *International Commission on Non-Ionizing Radiation Protection (ICNIRP)*, *International Electrotechnical Commission (IEC)*, entre muitas outras. A FDA, em 2003 (FDA, 2003) considera que a RM é um método de imagem que apresenta “*non-significant risk*” se os critérios e as normas de segurança, que estabelece, forem seguidas.

³Nestes exemplos optámos por utilizar referências que estão relacionadas com estudos de produção de fala.

2.1 Avanços tecnológicos na área da RM

Os contributos para a evolução desta modalidade imagiológica são tantos e tão variados que seria difícil incluí-los todos, com o destaque e pormenor que merecem, nesta revisão da literatura. Assim, como já referido, concentraremos a nossa atenção em algumas vertentes, nomeadamente as que estão associadas com as técnicas de aquisição rápida e que, de alguma forma, se relacionam com este trabalho que se enquadra no âmbito da produção de fala.

Estes avanços incluem, entre outros, o desenvolvimento e implementação de sequências de pulso ultra-rápidas, grandes melhorias a nível do desempenho (*performance*) dos gradientes e da intensidade do campo magnético, a utilização de técnicas de imagem paralela, beneficiando da tecnologia de antenas *phased-array*, o recurso a trajetórias alternativas (e.g., não cartesianas) de preenchimento do espaço K em conjunto com técnicas de subamostragem do mesmo e, mais recentemente, a utilização de técnicas que exploram a esparsidade do espaço K e correlações no tempo e espaço (Tsao *et al.*, 2003; Huettel *et al.*, 2004; Bresch *et al.*, 2008a; Lustig *et al.*, 2008; Larkman, 2009; Kim, 2010). De forma diferente, todos estes desenvolvimentos têm contribuído para acelerar o processo de aquisição e de reconstrução das imagens, permitindo diminuir os tempos de aquisição.

Apesar desta evolução, existirá sempre um compromisso entre a resolução temporal, a resolução espacial e a relação Sinal-Ruído (S/R) que terá que ser balanceado em função dos objetivos concretos de cada estudo e, mais importante, tendo em mente que a RM deverá continuar a ser um método de estudo e de investigação seguro, sendo necessário minimizar a ocorrência de efeitos biológicos.

2.1.1 Campo magnético principal - B_o

No que diz respeito à intensidade do campo magnético principal B_o , tem havido alterações significativas desde a introdução da RM na prática clínica. Em 1983, existiam magnetes com B_o entre 0.35 Tesla (T) a 0.5T. Rapidamente (meados da década de 80), surgiram *scanners* com 1T e 1.5T. Nos 20 anos seguintes, esta foi a intensidade dos magnetes que equipou a grande maioria dos equipamentos utilizados para fins clínicos (Theysohn *et al.*, 2008)⁴.

No início da década de 90, começam a surgir *scanners* com magnetes caracterizados por uma intensidade de campo bastante superior. Contudo, a aprovação pela *Food and Drug Administration* (FDA) (FDA, 2003) para a sua utilização na prática clínica só surgiu em 2003, sendo introduzidos os magnetes de 3T (já considerados na categoria de campos muito altos).

Magnetes com intensidade campo de 4.7T, 7T, 8T, 9.4T e 11.4T equipam, na atualidade, vários centros de investigação e o número de estudos que referem a utilização de campos com B_o entre 4-7T é já muito vasta (Stafford, 2004; Theysohn *et al.*, 2008). Caberá realçar que as recomendações relativas à utilização de RM em seres humanos condicionam a sua utilização a

⁴Na realidade, no presente, esta é a intensidade de campo que mais habitualmente se encontra nos equipamentos hospitalares.

campos de intensidade de 8T (Schenck, 2000; FDA, 2003; HPA, 2008)⁵.

De acordo com Theysohn *et al.* (2008), a crescente utilização de campos magnéticos de intensidade muito elevada terá que conduzir a uma maior consciencialização para os aspetos de segurança respeitantes aos voluntários, doentes e mesmo aos próprios operadores (e.g., técnicos ou investigadores).

É consensual que técnicas como a Ressonância Magnética funcional (RMf), a Espectroscopia por Ressonância Magnética (ERM), a imagem ponderada em difusão (*Diffusion Weighted Imaging* - DWI), a imagem baseada em suscetibilidade magnética (*Susceptibility Weighted Imaging* (SWI), entre outras, foram as que mais beneficiaram com a utilização de campos magnéticos de elevada intensidade, com um impacto muito grande, por exemplo, na área da neuro-imagem (Stafford, 2004).

De forma genérica, o principal interesse na utilização de campos de elevada intensidade está intrinsecamente relacionado com o aumento da S/R, da relação Contraste-Ruído (C/R) e com uma melhor resolução espectral (no contexto da ERM). No que diz respeito à S/R, esta em teoria, duplica num campo de 3T em relação a um campo de 1.5T, mas não se consegue um aproveitamento total desse aumento. Uma melhor S/R permitirá aquisições mais rápidas ou, em alternativa, melhor qualidade de imagem (e.g., melhor resolução espacial) para tempos de aquisição similares aos conseguidos com 1.5T. No entanto, como procuraremos evidenciar, a transposição das técnicas e parâmetros, quando se passa de um campo de 1.5T para um de 3T (ou superior), não é linear. Existem limitações tecnológicas, físicas e relacionadas com a segurança que condicionam uma completa exploração dos potenciais benefícios proporcionados pelos altos campos. De uma perspetiva tecnológica, refiram-se, por exemplo, os problemas relacionados com o aumento de artefactos (e.g., suscetibilidade magnética e de desvio químico), problemas de falta de homogeneidade de B_1 e alterações nos tempos de relaxamento T1 e T2 (veja-se Stafford, 2004 e Schick, 2005).

No tocante às questões de segurança, uma das limitações mais apontadas é o aumento dos valores do *Specific Absorption Rate* (SAR)⁶ nesses campos. Este facto constitui um dos grandes desafios no que diz respeito à utilização de campos magnéticos de elevada intensidade, uma vez que obriga ao estabelecimento de compromissos entre o SAR, a taxa de aquisição de imagem, a

⁵De acordo com as mesmas fontes, 8T será o máximo para adultos, crianças e bebés com idade superior a 1 mês. Para recém nascidos (idade < 1 mês) a intensidade de campo deverá ser inferior a 4T. Fora destes valores a FDA considera a RM como podendo apresentar “*significant risk*”. Note-se que, por essa razão, os estudos de investigação conduzidos com campos de intensidade mais elevada requerem autorizações específicas com aprovação de comissão de ética. Refira-se que os equipamentos entre 7-9.4 T são protótipos que não estão ainda disponíveis comercialmente. O primeiro estudo, efetuado com equipamento de 9.4 T, que envolveu doentes e que recebeu aprovação de comissão de ética decorreu, entre 2008–2012, no *Max Planck Institute for Biological Cybernetics* em Tübingen (Pohmann *et al.*, 2013)

⁶O SAR é definido como a potência de RF absorvida por unidade de massa tecido (em Watts/Kg) e está associado com o possível aumento de temperatura corporal do paciente na realização de um estudo de RM. Os valores máximos de SAR e aumento (máximo) de temperatura permitidos estão regulamentados para os vários modos de operação: modo normal de operação (adequado para todos os pacientes) – 0.5 °C ou 2 W/kg; modo controlado de operação – primeiro nível (supervisão médica) – 1.0°C ou 4 W/kg e segundo nível também designado como modo experimental – maior que 1.0°C e maior que 4 W/kg (requer aprovação pela comissão de ética). Estes valores dizem respeito à exposição média para corpo inteiro (média de 6 minutos), considerando condições de temperatura e humidade devidamente especificadas. Os limiares para algumas regiões do corpo específicas (e.g., extremidades, cabeça, tronco) e período de exposição são também definidos (Shellock, 2001). Veja-se também os documentos produzidos por diversas entidades reguladoras como a FDA (2003), IEC (2002) ou MHRA (2007).

cobertura anatômica e a qualidade de imagem (Stafford, 2004; Runge, 2007).

O SAR está relacionado com a energia dos pulsos de RF. Quanto maior a intensidade do campo magnético B_o , maior será a frequência de precessão dos núcleos de H (frequência de *Larmor* - ω_o) e também a sua energia. A equação 2.1 mostra a relação entre a intensidade do campo magnético e a frequência de precessão ω_o , onde (γ) é a constante giromagnética. Num campo de 3T, a frequência de precessão será de 128MHz (o dobro da frequência de precessão num campo de 1.5T), sendo o SAR e, consequentemente, a potência de RF depositada aproximadamente quatro vezes superior (Schick, 2005)⁷.

$$\omega_o = \gamma B_o / 2\pi \quad (2.1)$$

Para além da estreita dependência com intensidade do campo magnético, o SAR depende ainda do *Flip Angle* (FA), do *duty cycle* (ou tempo de operação), do número de cortes por Tempo de Repetição (TR) e do próprio tipo de bobines utilizadas para efetuar receção e emissão de sinal. Os valores de SAR dependem também da própria constituição do indivíduo (e.g., peso), da condutividade dos diferentes tecidos, entre outros fatores (Shellock, 2001; Elster e Burdette, 2001; Bernestein *et al.*, 2004). O aumento de temperatura no indivíduo, decorrente da realização do estudo, depende do SAR, mas também de outras condições, como a temperatura ambiente (da sala), do sistema termo-regulador do indivíduo e da região corporal em estudo.

O aumento do SAR é particularmente relevante quando se pretendem utilizar sequências Eco de Spin (ES) ultra-rápidas, como as *Fast ou Turbo Spin Echo* (FSE ou TSE) ou *Single-Shot FSE* (SS-FSE)⁸, efetuar saturação química da gordura ou recorrer a sequências com transferência de magnetização em campos de elevada intensidade (Stafford, 2004).

Quando se utilizam sequências Eco de Gradiente (EG), os problemas com o SAR são, usualmente, menos críticos, uma vez que os pulsos de excitação são menos energéticos com recurso a FA parcial (menor que 90°). No entanto, em alguns estudos angiográficos com administração de meios de contraste (*Contrast Enhancement Magnetic Resonance Angiography* - CE MRA), onde é requerida uma resolução temporal elevada e se recorre a sequências EG (*Steady-State Free Precession* - SSFP), com equilíbrio perfeito dos gradientes (e.g., sequência TrueFISP), facilmente são atingidos, nos campos de 3T, os limites impostos para o SAR. Isso deve-se ao facto de a maximização do sinal passar pela utilização de FA relativamente elevado com TR extremamente curto (Schick, 2005).

A utilização de FA mais pequeno ou a utilização de imagem paralela (muito relevante nos campos de elevada intensidade) podem contribuir para uma diminuição do SAR ou para encurtar o tempo de aquisição, mas podem conduzir a alterações no contraste (situação nem sempre desejável) ou a uma diminuição da S/R. A introdução de sequências TSE que utilizam técnica de FA variável, ao longo da cadeia de ecos, permite uma redução significativa do SAR (Hennig *et al.*, 2003; Runge,

⁷Para aquisição de imagens com características semelhantes, o SAR será dez vezes superior num campo de 9.4T comparativamente a um campo de 3T (Pohmann *et al.*, 2013).

⁸A introdução de vários pulsos de refocalização de 180°, altamente energéticos, para acelerar a aquisição, vai conduzir a um aumento substancial do SAR.

2007; Pohmann *et al.*, 2013), mantendo uma S/R elevada e contraste (T2) adequado (e.g., técnicas designadas como *Hyperecho* TSE e TRAPS - *Smooth Transitions between Pseudo Steady States*).

De acordo com Stafford (2004), o processo de otimização a utilizar nos campos elevados passará, por exemplo, por alterações na concepção das sequências de pulso, do próprio *design* dos pulsos de RF e das bobines de RF, o recurso a novas técnicas de aquisição e de amostragem do espaço K, que contribuam para ultrapassar os constrangimentos impostos pelos limites fisiológicos⁹.

2.1.2 Gradientes de campo magnético

Os gradientes de campo magnético são uma componente muito importante de um sistema de RM. As bobines de gradiente criam uma variação linear do campo magnético ao longo de uma direção espacial. Esta variação do campo magnético sobrepõe-se a B_0 , aumentando ou diminuindo a sua intensidade, ao longo dessa direção, permitindo, desta forma, efetuar a selecção de corte e, dentro deste, a codificação em fase e em frequência (Brown e Semelka, 2010). Na figura 2.1, pode observar-se a variação do campo magnético com a distância, ao longo de uma direção espacial.

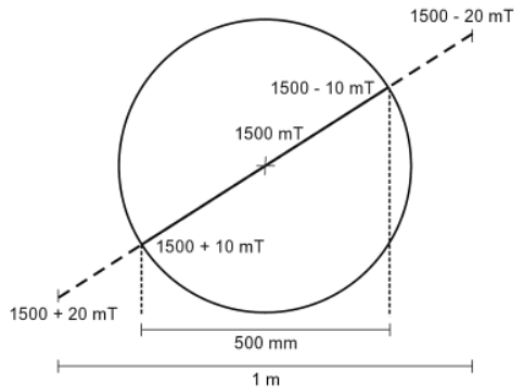


Figura 2.1: Um gradiente é uma distorção linear do campo magnético principal, centrado com o isocentro do magneto. A sobreposição de um gradiente com 40 mT/m com um B_0 de 1.5 T produzirá uma variação de campo magnético entre 1480 a 1520 mT . Para uma distância de 500 mm a variação será de 1490 a 1510 mT . Fonte da imagem: Brown e Semelka (2010).

Também a este nível têm existido melhorias muito significativas, ao longo dos anos, viabilizando a implementação de técnicas (e.g., *Echo-Planar Imaging* ou Imagem Eco Planar (EPI)), que apesar de descritas (teoricamente) há alguns anos, exigem um desempenho ao nível dos gradientes que só mais recentemente veio a ser conseguido. Note-se que a técnica EPI foi proposta por Mansfield em 1977.

Os equipamentos mais antigos eram equipados com gradientes com amplitude máxima de 10 a 15 mT/m . Atualmente, existem sistemas com gradientes que permitem uma amplitude máxima de 50 a 60 mT/m ou superior¹⁰. A amplitude máxima define-se em mT/m e é muito

⁹Para um leitor mais interessado nesta área sugere-se a leitura de um artigo bastante recente, já aqui referenciado, uma vez que este proporciona uma visão global dos desafios colocados na utilização de campos de elevada intensidade, nas estratégias para os ultrapassar e no seu potencial futuro (veja-se Pohmann *et al.*, 2013).

¹⁰No que diz respeito à amplitude máxima, é frequente ler-se que quanto maior esta for melhor (Liang e Lauterbur,

importante no que diz respeito à resolução espacial, uma vez que condiciona a espessura mínima de corte utilizável, para um determinado *Field of View* ou campo de visão (FOV), sem que se alterem outros parâmetros (Brown e Semelka, 2010).

Apesar de ser das especificações de que mais se fala, não é só a amplitude máxima dos gradientes que interessa considerar para a sua caracterização. Na realidade, para muitas aplicações (e.g., *time-resolved MRI*), este não é o parâmetro com maior impacto no seu desempenho. Assim, para além da sua amplitude máxima, interessa considerar o tempo de subida (*rise time*), a taxa de variação (*slew rate*) e o *duty cycle* ou tempo de operação (Liang e Lauterbur, 2000; Lafuente, 2000; Elster e Burdette, 2001; Brix *et al.*, 2008; Brown e Semelka, 2010). Para Brown e Semelka (2010), haverá ainda que considerar, como ponto importante do seu desempenho, a forma como é efetuada a compensação das correntes *eddy*¹¹.

Após serem percorridos por uma corrente elétrica, os gradientes não atingem instantaneamente a sua máxima intensidade. O tempo de subida do gradiente é, por definição, o período de tempo que este demora a atingir a sua máxima amplitude. O tempo de subida determina a taxa de variação do gradiente que é calculada dividindo a amplitude máxima pelo tempo de subida. O tempo de subida usualmente é expresso em *ms* ou μs e a taxa de variação em $T/m/s$.

Estas duas características dos gradientes são das especificações mais importantes a ter em consideração, quando se pretende obter elevada resolução temporal como no caso de aquisições em tempo real, em estudos cardíacos e angiográficos ou em estudos funcionais, quando se utiliza técnica EPI. Tempos de subida muito curtos (elevada taxa de variação) só são atingidos com os gradientes de alto desempenho. Com gradientes desta gama, é possível encurtar o Tempo de Eco (TE) e o TR, diminuir o espaçamento entre ecos (*Echo spacing* - ESP) quando se utilizam sequências FSE ou EPI e, em consequência, encurtar significativamente os tempos de aquisição (Elster e Burdette, 2001; Brown e Semelka, 2010).

Para que se tenha uma noção do que tem sido a evolução neste domínio, podemos referir que, há uns anos atrás, o mais usual seria a existência de *scanners* equipados com gradientes com taxa de variação à volta de $20 T/m/s$ e tempos de subida de $800 \mu s$ e, na atualidade, são frequentes sistemas que permitem atingir taxa de variação da ordem dos $200 T/m/s$, com tempos de subida entre 100 a $200 \mu s$.

O tempo de operação do gradiente é importante, na medida em que determina com que rapidez o amplificador do gradiente consegue responder aos requisitos da sequência de pulso. Nos sistemas mais avançados é possível ter tempo de ocupação de 100% para a máxima amplitude dos gradientes. O tempo de operação influencia o TE e o espaçamento entre ecos (Brown e Semelka, 2010).

Dois sistemas de gradientes podem apresentar a mesma amplitude máxima, e desempenhos completamente diferentes, no que diz respeito à sua rapidez. Na tabela 2.1, apresenta-

2000; Elster e Burdette, 2001). Por vezes, para além da amplitude máxima considera-se a amplitude efetiva dos gradientes que corresponde ao vetor soma dos gradientes nas três direções espaciais (Brown e Semelka, 2010).

¹¹As correntes *eddy* são correntes elétricas induzidos num condutor por um campo magnético variável. A variação muito rápida dos gradientes (*on/off*) origina essas correntes (no próprio sistema) podendo dar origem a artefactos graves (e.g., distorções) na imagem. Estas correntes podem também ser induzidas nos pacientes e nos cabos de RF (Elster e Burdette, 2001; Bernestein *et al.*, 2004).

se, como exemplo, o impacto dos vários parâmetros em três tipos de estudos (angiografia renal, colangio-RM e perfusão cardíaca) realizados por rotina na maior parte das instituições. A amplitude máxima, FOV e matriz para cada uma das situações apresentadas são idênticos.

Tabela 2.1: *Performance dos gradientes - Implicações práticas.* Fonte da imagem: Lafuente (2000).

Amplitude	20 mT/m	20 mT/m	20 mT/m
Tempo de subida	800 μs	400 μs	240 μs
Taxa de variação	25 $T/m/s$	50 $T/m/s$	125 $T/m/s$
3D GE MRA renal			
Tempo aquisição (TA) mín	16 s	13 s	9 s
HASTE - CPRM			
Tempo de Eco (TE) mín	4.8 ms	3.9 ms	3.2 ms
Turbo - FLASH			
Perfusão Cardíaca			
Tempo de Repetição (TR) mín	3.3 ms	2.4 ms	2.0 ms

Tal como acontece com os constrangimentos relativos ao SAR, também a evolução a nível do desempenho dos gradientes é muito condicionada por questões fisiológicas e de segurança, mais do que pelo *hardware* em si (Shellock, 2001; Larkman, 2009).

O ruído acústico associado às aquisições de RM¹² é um dos problemas, relacionados com a segurança, que tem que ser tido em consideração. Com variações bruscas dos gradientes, necessárias quando se utilizam sequências de pulso muito rápidas (e.g., *Single-Shot* - EPI) podem ser atingidos índices de ruído significativamente elevados (130 dB), valores estes bem acima dos limites recomendados¹³. No sentido de diminuir estes valores, têm vindo a ser sido implementadas algumas técnicas para controlo do ruído, como *active noise cancelation* ou *anti-noise*, desenvolvimento de gradientes com *design* adequado que permitem atenuar esses valores elevados (e.g., *quiet MRI gradient coils*) e alterações a nível das sequências de pulso (e.g., *STEAM-Burst*) (Shellock, 2001). No entanto, estas técnicas não estão ainda amplamente incorporadas na maior parte dos equipamentos.

Apesar dos índices de ruído elevados poderem constituir um problema, a utilização de tampões e/ou auscultadores durante as aquisições, obrigatória a partir de 99 $dB(A)$, contribui para atenuar estes valores¹⁴.

A maior condicionante fisiológica relacionada com o desempenho dos gradientes, nomeadamente a sua rápida variação imposta por algumas sequências de pulso rápidas e ultra-rápidas,

¹²O ruído acústico é originado pelas forças de *Lorenz* induzidas nas bobines de gradiente quando percorridas pela corrente elétrica. A variação da corrente induz deformação nas bobines fazendo-as vibrar. A vibração transmite-se a outras componentes do *scanner* e através do ar ao paciente ou voluntário. As correntes de *Lorenz* são proporcionais à intensidade da corrente que atravessa as bobines (consequentemente, à amplitude do gradiente) e, também à intensidade do campo magnético. Sequências com parâmetros para obtenção de alta resolução (FOV pequeno, espessura corte baixa e matrizes elevadas), a utilização de TR, TE e ESP curtos e aquisições em volume conduzem a índices de ruído mais elevados (HPA, 2008).

¹³Os limites estabelecidos pela FDA (2003) e IEC (2002) referem que estes valores não devem exceder os 99 $dB(A)$.

¹⁴A utilização conjunta de auscultadores e tampões pode diminuir entre 10 a 30 $dB(A)$ os índices de ruído minimizando a perda temporária de audição e prevenir danos mais graves a nível do ouvido. Contribui também para minimizar o desconforto e o *stress* originados pela exposição ao ruído. Por essa razão a sua utilização é recomendada mesmo para valores abaixo dos 99 $dB(A)$ (IEC, 2002; MHR, 2007; HPA, 2008).

é a possibilidade de ocorrer Estimulação Nervosa Periférica (ENP)¹⁵ ou, mais grave, fibrilhação ventricular. Esta última situação será muito pouco provável de ocorrer em pacientes ou voluntários sem patologia cardíaca, no contexto dos protocolos utilizados por rotina na prática clínica (Schaefer *et al.*, 2000; Shellock, 2001).

No entanto, a ENP pode tornar-se uma situação de tal forma desconfortável ou mesmo intolerável para o doente que pode constituir um motivo para interromper a aquisição ou mesmo terminá-la. Quando a aquisição cessa, esta estimulação nervosa termina, não estando, até ao momento, associada a outros efeitos nocivos. De forma semelhante ao que acontece para o SAR, estão definidos os limiares para a ENP, para os diferentes modos de operação¹⁶.

No que toca a este aspeto, a verdade é que não é ainda muito clara, uma vez que só agora se começam a efetuar mais estudos relativos ao assunto, qual a relação entre os limiares para ocorrência de ENP - tanto que é variável com os indivíduos - com as dimensões das bobines de gradiente e com os parâmetros relacionados com o seu desempenho. Há indícios que bobines mais pequenas serão mais vantajosas do que as de grandes dimensões e que a própria posição do paciente em relação à bobine é um fator importante a ter em consideração. Já há alguns desenvolvimentos no sentido de conceber gradientes com forma de onda otimizada, por forma a conseguir obter melhor desempenho com minimização da ocorrência de ENP (Rutt, 2003).

2.1.3 Tecnologia de antenas

No que diz respeito à tecnologia de antenas, o maior avanço nesta área foi conseguido com a introdução, no início da década de 90, das chamadas bobines *phased array* ou também chamadas *multicoil arrays* (Roemer *et al.*, 1990).

"The NMR phased array offers the signal-to-noise ratio (SNR) and resolution of a small surface coil over fields-of-view (FOV) normally associated with body imaging with no increase in imaging time." (Roemer *et al.*, 1990)

As antenas *phased array* melhoram a S/R através da redução da dimensão da antena (*sensitive volume*), reduzindo a amplitude do ruído detetado. Cada bobine recetora recebe sinal (ver figura 2.2) de forma independente, através de um canal próprio, mas, quando conjugadas (*array*) com sobreposição ligeira dos seus perfis de sensibilidade, conseguem uma cobertura (FOV) equivalente à de uma antena de maior dimensão. Desta forma, o sinal captado é sensivelmente o mesmo que o de uma antena maior, mas o ruído é grandemente reduzido, aumentando a S/R. A

¹⁵A variação muito rápida dos gradientes de campo magnético exigidos para aquisições muito rápidas dos dados podem induzir correntes elétricas nos tecidos (principalmente a nível do sistema nervoso periférico). Pensa-se que este é o mecanismo subjacente à indução de ENP (Rutt, 2003).

¹⁶Os valores apresentados em ICNIRP (2004), MHRA (2007) e HPA (2008) referem que o sistema de gradientes deve operar de forma a não ser excedido 80% do limiar médio de perceção para ocorrência de ENP para o modo normal de operação; para o modo controlado, não deverá exceder 100% desse limiar médio; para o modo experimental não é definido um valor, referindo-se apenas que poderá ser superior ao do modo controlado (acima de 100%) tendo que haver especial controlo para evitar estimulação cardíaca. É consensual que sempre que a ENP se torne intolerável para um paciente ou voluntário esta situação deverá ser evitada. A informação utilizada para estabelecimento destes limiares foi essencialmente baseada nos resultados dos estudos efetuados por Schaefer *et al.* (2000) e por Nyenhuis *et al.* 2001.

utilização de algoritmos de reconstrução adequados permite obter uma combinação otimizada de sinal e ruído na imagem final, obtida a partir dos contributos de cada um dos elementos (Bernstein *et al.*, 2004; Fujita *et al.*, 2006; Brix *et al.*, 2008; Deshmane *et al.*, 2012).

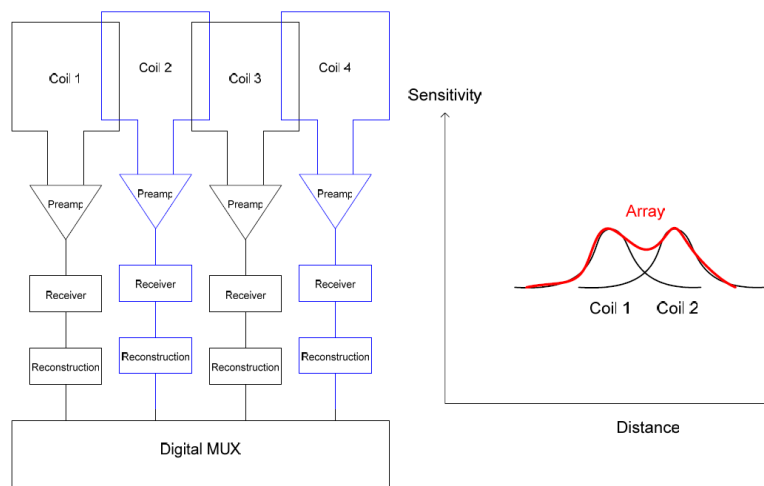


Figura 2.2: Representação esquemática de uma bobine *phased array* de 4 canais (à esquerda). Cada bobine tem o seu perfil de sensibilidade relativo a B_1 que é depois conjugado com o de outras bobines. Fonte da imagem: Fujita *et al.* (2006).

Os últimos anos foram dominados por grandes desenvolvimentos nesta tecnologia, nomeadamente através do aumento do número de canais e da possibilidade de agrupar um grande n° de bobines. Tornou-se possível obter cobertura anatômica de todo o corpo (ver figura 2.3) com elevada sensibilidade, sem necessidade de reposicionar ou levantar o paciente, o que se torna muito vantajoso para alguns estudos (Schick, 2005; Fujita *et al.*, 2006; Brix *et al.*, 2008).

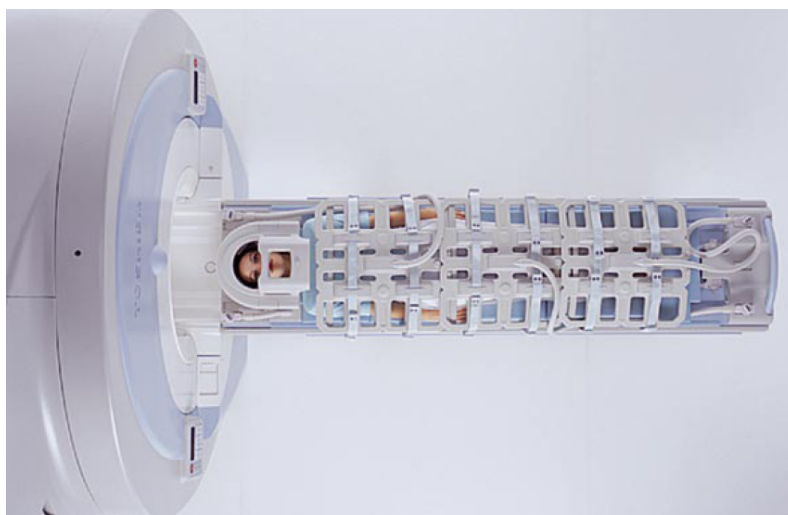


Figura 2.3: Conjunto de *arrays* para cobertura anatômica de todo o corpo (Exelart Vantage™, Toshiba). Fonte da imagem: Brix *et al.* (2008).

Para a maior parte das aplicações clínicas, o número de canais independentes (elementos) das bobines varia entre os quatro e trinta e dois (Brix *et al.*, 2008; Deshmane *et al.*, 2012). Como

os *arrays* podem ser dispostos em arranjos planares ou circulares, surgiram bobines mais dedicadas às especificidades dos diferentes estudos e partes anatómicas. Por exemplo, para estudo da coluna vertebral, as antenas apresentam um arranjo linear, enquanto que as de crânio têm arranjo circular. Usualmente, as bobines são concebidas de forma a permitir (com técnicas de imagem paralela) obter maior aceleração ao longo da direção de codificação da fase (Deshmane *et al.*, 2012).

Uma vez que a tecnologia de captação de sinal é uma componente extremamente importante nos sistemas de RM, algumas equipas, que trabalham na área da produção de fala, têm vindo a apostar no desenvolvimento de antenas dedicadas, permitindo melhorar a qualidade de imagem ou obter maior rapidez (veja-se, por exemplo, Kim, 2010 ou Zhu *et al.*, 2013a).

2.1.4 Técnicas de aquisição rápida

2.1.4.1 Sequências de pulso rápidas e ultra-rápidas

As sequências de pulso são as ferramentas através das quais é possível obter sinais, espacialmente referenciados, que permitem a posterior reconstrução das imagens de RM. São, no fundo, os programas que controlam todos os aspectos de *hardware* necessários para adquirir os dados da forma desejada (e.g., pulsos de RF, gradientes). O tempo entre os pulsos, a amplitude e a forma do gradiente, entre outros, controlam o sinal de RM, condicionando as características das imagens (Brown e Semelka, 2010). De acordo com Bernestein *et al.* (2004), as sequências de pulso são uma parte vital do processo de obtenção de imagem através da RM, contribuindo para a sua grande versatilidade, quando comparada à de outras modalidades de imagem.

A alteração de alguns parâmetros das sequências de pulso é condicionada pela forma como o fabricante implementou a sequência (e.g., TR mínimo) ou pelas características técnicas do próprio *scanner*. Todavia, muitos parâmetros podem ser alterados pelo operador (e.g., TR, TE, FA, matriz), permitindo manipular o contraste nas imagens, a resolução espacial e o tempo de aquisição em função dos objetivos ou requisitos de cada estudo.

A variedade de sequências disponível nos *scanners* mais recentes é imensa e a quantidade de acrónimos para designar essas sequências é ainda maior, podendo tornar-se assustador, mesmo para quem trabalha diariamente com RM e é conhecedor de literatura específica na área. Isto deve-se, em grande parte, ao facto de cada fabricante utilizar um acrónimo diferente para sequências que serão, na sua base física, muito semelhantes (Brown e Semelka, 2010). A utilização destes acrónimos nesta revisão será por isso inevitável. Contudo, uma revisão exaustiva de todo o tipo de sequências de pulso e suas variantes não faria, no contexto deste trabalho, muito sentido. Assim, faremos menção apenas a sequências mais rápidas, referindo, em particular, as que virão a ser utilizadas no estudo ou referenciadas ao longo dos restantes capítulos. Informação mais detalhada relativa a este tópico pode ser encontrada em livros e artigos de revisão como: Haacke *et al.* (1999); Elster e Burdette (2001); Bernestein *et al.* (2004); Boyle *et al.* (2006); Chavhan *et al.* (2008); Brix *et al.* (2008).

As sequências de pulso consideradas rápidas ou ultra-rápidas têm capacidade de adquirir imagens em escassos segundos, ou em frações de segundo, dependendo dos requisitos necessários

para cada estudo ou situação. Podem pertencer à classe de sequências SE (neste caso, denominam-se FSE ou TSE) ou à classe de sequências EG (neste caso, designam-se como sequências EG rápidas).

Nas sequências TSE ou FSE, após a aplicação de um pulso de excitação de 90° , segue-se a aplicação de um conjunto de pulsos de refocalização de 180° . Os ecos resultantes, codificados em fase e frequência, permitirão preencher, não uma linha do espaço K, mas um número de linhas igual ao de pulsos de 180° utilizados. Desta forma, a aquisição pode ser substancialmente acelerada. O conjunto de pulsos de 180° é designado por cadeia de ecos (*Echo Train Length- ETL*) ou *factor turbo* (Elster e Burdette, 2001; Bernestein *et al.*, 2004).

Após a aplicação do pulso de excitação, a magnetização transversal (decaimento T2) decresce exponencialmente diminuindo a Intensidade de Sinal (IS). Este efeito, restringe o número máximo de ETL que pode ser adquirido, condicionando a dimensão da matriz (na direção de codificação da fase) e, conseqüentemente, a resolução espacial para um dado FOV. Por exemplo, com ESP de 4 ms (valor típico), o número de ecos é limitado a 100. Estes ecos podem ser usados para preencher o espaço K numa matriz com 256×96 elementos (96 passos de codificação da fase) ou, uma matriz de 256×192 elementos, se utilizada em conjunto com técnica *Half-Fourier* (HF) (Dietrich, 2007a). Se o ESP for diminuído, o que pode ser conseguido aumentando a largura de banda (*Bandwidth - BW*) ou diminuindo a duração dos pulsos de RF, pode-se aumentar a rapidez de aquisição mas, em contrapartida, há uma diminuição da S/R e aumento do SAR (Dietrich, 2007a).

O contraste nas sequências TSE é determinado pelo TE efetivo (TE_{ef}). Este representa o tempo entre a aplicação do pulso de RF e a leitura das linhas centrais do espaço K (baixas frequências espaciais). Quanto maior for o TE_{ef} , melhor será a ponderação em T2, mas mais pronunciados serão os artefactos de *blurring*¹⁷, característicos desta sequência, que podem contribuir para a degradação da resolução espacial (Elster e Burdette, 2001; Carroll *et al.*, 2006; Brix *et al.*, 2008).

A grande desvantagem das sequências TSE é o aumento do SAR, devido à maior energia dos pulsos de 180° . Também se verificam algumas alterações de contraste, particularmente a nível do sinal da gordura (apresenta-se hiperintensa devido a interações intramoleculares - efeito de *J-Coupling*) e, devido a efeitos de transferência de magnetização, pode ocorrer diminuição da IS em tecidos (estacionários) ricos em proteínas, como a substância branca e substância cinzenta (Elster e Burdette, 2001; Carroll *et al.*, 2006).

Nas sequências EG, ao invés de se formar um eco de *spin*, forma-se um eco de gradiente, após aplicação de um único pulso de excitação (com FA parcial), seguida pela atuação de um gradiente (bipolar) de reversão. Depois da aplicação do pulso de RF, o lobo negativo do gradiente de codificação de frequência (*pre-phasing gradient*) induz dispersão de fase nos *spins* em precessão (desfasamento). Com a reversão do gradiente (polaridade oposta - *readout gradient*),

¹⁷O artefacto de *blurring* (desfocagem) é atribuído à perda de sinal (devido ao decaimento T2 dos tecidos) que ocorre ao longo da cadeia de ecos. As conseqüências para a imagem serão diferentes, em função dos tecidos a ser estudados, uma vez que estes têm tempos de relaxamento T2 diferentes. Para um determinado ETL, o *blurring* será menor com ESP curto, todavia, como já referido, há implicações a nível da S/R (Carroll *et al.*, 2006).

há refocalização destes *spins* e origina-se um eco de gradiente. Como não são utilizados pulsos de refocalização de 180° , como acontece nas sequência ES, só contribuem para o sinal os *spins* que foram desfasados e refocalizados pela ação do próprio gradiente. O decaimento, dito $T2^*$, é muito rápido, uma vez que para este contribui o desfasamento devido a efeitos de suscetibilidade magnética, desvio químico, não homogeneidade do campo, entre outros¹⁸, que não são, neste caso, compensados. Estas sequências são, por esse motivo, muito sensíveis a artefactos de suscetibilidade magnética e de desvio químico. Uma das grandes vantagens é que permitem a utilização de TR muito curtos, permitindo diminuir consideravelmente o tempo de aquisição. Para além disso, abriram caminho à exploração de novas formas de contraste (Elster e Burdette, 2001; Bernstein *et al.*, 2004; Brown e Semelka, 2010).

A forma como se podem dividir as sequências EG nem sempre é muito consensual na literatura (veja-se Elster e Burdette, 2001 ou Scheffler e Lehnhardt, 2003). Pode-se, no entanto, considerar dois grandes grupos: 1) sequências EG -*Spoiled*, também designadas, por alguns autores, sequências EG *Steady-State* incoerente e 2) sequências EG *Steady-State* coerente¹⁹.

Nas primeiras, antes da aplicação de um novo pulso de excitação, é efetuado, através do gradiente de seleção de corte ou através de pulsos de RF, um *spoiling* que conduz à destruição da magnetização transversa residual. No início do pulso seguinte, esta componente da magnetização é nula ($M_T = 0$), não contribuindo para o sinal (Elster e Burdette, 2001; Chavhan *et al.*, 2008).

Na figura 2.4, apresenta-se um diagrama de pulso genérico relativo às sequências EG -*Spoiled*. São exemplos de sequências EG - *Spoiled* a sequência FLASH (*Fast Low angle Shot*) e a sequência TurboFLASH.

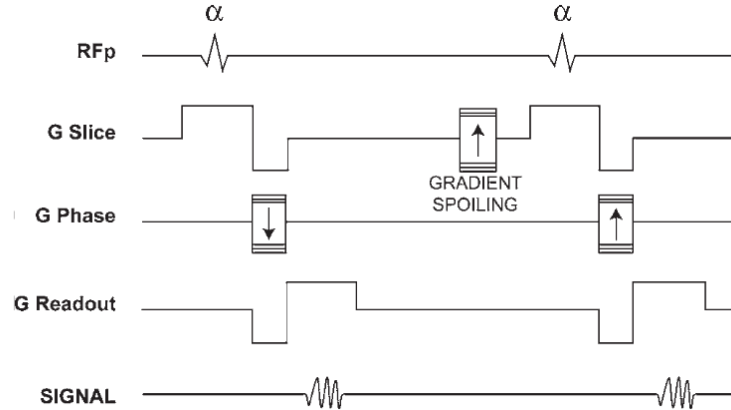


Figura 2.4: Diagrama de pulsos de uma sequência EG - *Spoiled*. Fonte da imagem: Chavhan *et al.* (2008).

A sequência **FLASH** foi introduzida por Haase *et al.*, em 1986, permitindo uma redução

¹⁸No seu conjunto estes efeitos são designados como *off-resonance effects* (Bernstein *et al.*, 2004; Chen e Meyer, 2006).

¹⁹Na verdade, alguns autores consideram que todas as sequências EG rápidas são sequências *Steady-State* podendo, no entanto, ser estabelecido diferente tipo de *Steady-State* em função do padrão de comutação dos gradientes e, em consequência, obter diferente tipo de contraste (Scheffler e Lehnhardt, 2003; Chavhan *et al.*, 2008). Outros autores, consideram que as verdadeiras sequências *Steady-State* são aquelas em que se cria um equilíbrio das componentes transversal e longitudinal da magnetização o que não acontece nas sequências EG -*Spoiled* onde o *Steady-State* apenas diz respeito à componente longitudinal da magnetização (Elster e Burdette, 2001).

drástica (cerca de $100\times$) dos tempos de aquisição, sem sacrifício da resolução espacial (e.g., 6 s com matriz 256×256) através da utilização de FA pequeno e TR muito curto (Haase *et al.*, 1986). A sua implementação veio viabilizar, por exemplo, a aquisição de vários cortes numa apneia (fundamental para estudos abdominais), aquisições dinâmicas após administração de gadolínio e aquisições 3D com tempo de aquisição razoável (dependendo da resolução requerida).

O contraste nesta sequência depende essencialmente do tempo de relaxamento T1, uma vez que a coerência transversal é destruída. A ponderação das imagens é, por isso, predominantemente em T1²⁰. Na atualidade, a utilização da sequência FLASH é rotina em variados estudos (e.g., do foro músculo-esquelético ou em neurorradiologia), para além das aplicações em estudos abdominais, pélvicos e cardíacos (Elster e Burdette, 2001; Carroll *et al.*, 2006; Brix *et al.*, 2008).

A sequência **TurboFLASH**²¹ ou *snapshot* FLASH foi também proposta por Haase *et al.* em 1989. Constituiu um desenvolvimento da sequência FLASH com o intuito de encurtar, ainda mais, os tempos de aquisição, tornando-os comparáveis aos conseguidos com a aquisição EPI (Haase, 1990). Valores de TR menores que 10 ms são usuais com o TurboFLASH, permitindo tempos de aquisição abaixo do segundo.

A utilização de TR muito curtos (e.g., $TR < T2^*$) tem, no entanto, algumas consequências para a imagem, uma vez que as componentes longitudinal e transversal da magnetização não atingem o equilíbrio e variam ao longo da aquisição. Como consequência, há violação da simetria do espaço K com uma sub ou sobre-amostragem de algumas frequências espaciais, dando origem a um ligeiro *blurring* da imagem, que é característico desta sequência (Elster e Burdette, 2001; Brix *et al.*, 2008). Para manter uma S/R aceitável, inerentemente baixa nesta sequência, e minimizar os efeitos de saturação ao longo da sequência de pulsos (com TR tão baixo) terá que ser utilizado FA muito pequeno. Desta forma, as imagens apresentam baixo contraste, essencialmente baseado na densidade protónica dos tecidos. Mesmo com estas características, para além da sua usual utilização em sequências de localização (*localizer*), esta sequência é muito útil (devido à sua rapidez) em estudos dinâmicos, após administração de gadolínio, permitindo boa quantificação da captação de contraste. Uma vez que o SAR é relativamente baixo nesta sequência, devido à utilização de FA pequeno, pode ser uma grande vantagem para aquisição de imagem em campos de intensidade elevada (Elster e Burdette, 2001).

Se à sequência for adicionada uma fase com preparação da magnetização, é possível manipular o contraste, permitindo adquirir imagens clinicamente mais úteis. Nesta fase preparatória da sequência, pode ser aplicado um pulso, não seletivo, de 180° que permite otimizar o contraste da imagem (e.g., uma melhor ponderação em T1). Na fase de aquisição, a matriz no espaço K é preenchida de uma só vez (*single-shot*) ou de forma segmentada (apenas algumas linhas -

²⁰Note-se que, nas sequências EG, o FA tem um papel muito relevante na ponderação que se pode obter nas imagens. Apesar do contraste no FLASH ser predominantemente T1, este pode ser manipulado permitindo obter imagens com ponderação em densidade protónica ou T2*. Por exemplo, a utilização de TR e TE longos com FA pequenos permite obter imagens ponderadas em T2*. No que toca à obtenção de imagens ponderadas em T1, quando maior for o FA melhor será a ponderação T1 mas maior será o ruído. Desta forma, o mais usual é a utilização de FA à volta de 50° , que permite obter boa ponderação em T1 e melhor qualidade da imagem (Elster e Burdette, 2001).

²¹Esta é a designação utilizada pela Siemens, na GE chama-se *Fast SPOiled GRASS* (SPGR), na Philips T2-Turbo Field Echo (TFE) e na Toshiba *Fast Field Echo* (FFE).

multi-shot) (Elster e Burdette, 2001).

A sequência TurboFLASH (com preparação da magnetização) é muito utilizada no âmbito de estudos cardíacos, por exemplo, em estudos de perfusão do miocárdio. O Tempo de Inversão (TI) pode ser ajustado de forma a que o sinal do miocárdio saudável seja nulo (*Miocardial signal nulling*). Para além da sua aplicação em estudos cardíacos (e.g., Fenchel *et al.*, 2004), de forma semelhante, o TurboFLASH tem sido muito utilizado em estudos de perfusão hepática e renal (e.g., Michaely *et al.*, 2007).

Se for aplicado um esquema de pulsos preparatório um pouco diferente (90° - 180° - 90°), consegue-se obter imagens ponderadas em T2. Esta técnica é conhecida como *Driven Equilibrium* (Elster e Burdette, 2001). Quando adquirida em 3D, com a utilização do pulso de inversão, designa-se como 3D MP-RAGE (*Magnetization Prepared - Rapid Acquired Gradient Echo*). Esta sequência proporciona a aquisição de imagens de elevada resolução (e.g., 3D de todo o crânio) com boa ponderação em T1 em tempos de aquisição aceitáveis (Brix *et al.*, 2008).

A sequência **VIBE** (*Volume Interpolated Breath-hold Examination*), proposta por Rofsky *et al.* em 1999, pertence também à classe das sequências 3D EG *Spoiled* ultra-rápidas²². Esta sequência permite obter volumes de elevada resolução, com ponderação em T1, em períodos de tempo muito curtos. A otimização desta sequência para obtenção de imagem de forma rápida e com elevada resolução espacial passa pela utilização de interpolação (*zero-filling* na direção de seleção de corte) e amostragem assimétrica do espaço K (Rofsky *et al.*, 1999).

Originalmente, a sequência VIBE foi pensada para obtenção de imagem a nível abdominal, permitindo uma cobertura anatômica extensa, em escassos segundos, durante uma apneia. No entanto, a sua aplicação na obtenção de imagens em outras áreas do corpo depressa ocorreu, com excelentes resultados (e.g., crânio, tórax, joelho). Se o tempo de aquisição não for o fator mais crítico, este pode ser aumentado, permitindo, em alternativa, melhorar a resolução espacial com *voxel* isotrópico (Dale *et al.*, 2006).

Sequências EG Steady-State

Nas sequências *Steady-State* (coerente), a magnetização transversa residual não é destruída, sendo aproveitada para o pulso seguinte, contribuindo para um aumento de sinal. Como, geralmente, são utilizados TR muito curtos e pouco espaçados no tempo, ao fim de um pequeno período de tempo, cria-se um estado de equilíbrio estável entre a magnetização no plano transversal e a magnetização no plano longitudinal²³.

Estabelecido este estado de equilíbrio, formam-se dois tipos de sinal designados por: 1) sinal de pós-excitação ($S+$), que resulta do decaimento (*Free Induction Decay*- FID) dos pulsos de

²²VIBE é o acrónimo utilizado pela Siemens. Na GE a sequência equivalente designa-se LAVA (Liver Acquisition with Volume Acceleration), na Philips é designada como THRIVE (T1 High Resolution Isotropic Volume Excitation)

²³O *Steady-State* entre a magnetização longitudinal e transversal é atingido se o TR for muito mais curto que o tempo de relaxamento T2 dos tecidos. Desta forma, não há tempo para que a magnetização transversal decaia completamente antes da aplicação do pulso seguinte. Assim, esta magnetização residual vai ser refocalizada no plano longitudinal quando da aplicação do pulso de RF seguinte. Ao mesmo tempo, a componente da magnetização longitudinal é projetada sobre o plano transversal de forma que, ao fim de um determinado número de TR's, é atingido um equilíbrio entre as duas componentes da magnetização (Elster e Burdette, 2001; Chavhan *et al.*, 2008).

RF mais recentes e 2) outro sinal que resulta de um eco reformatado, que ocorre antes da excitação ($S-$). O primeiro conduz a uma ponderação mista em $T1$ e $T2^*$ e o segundo ($S-$) é fortemente ponderado em $T2$, com efeito $T2^*$ desprezível (Chavhan *et al.*, 2008).

Em função do tipo de sinal que é utilizado para formar a imagem também assim se dividem estas sequências: a) quando é lido apenas o sinal ($S+$) ou FID (e.g., *Fast Imaging in Steady Precession* ou FISP na Siemens, *Gradient-Recalled Acquisition in the Steady-State* ou GRASS na GE medical); b) quando é utilizado o eco de spin ($S-$) e, finalmente c) sequências *Balanced Steady-State Free Precession* (b-SSFP) em que ambas as componentes do sinal ($S+$) e ($S-$) contribuem para a formação da imagem e existe um equilíbrio perfeito dos gradientes (Chavhan *et al.*, 2008). A figura 2.5 apresenta o diagrama de pulsos de uma sequência b-SSFP.

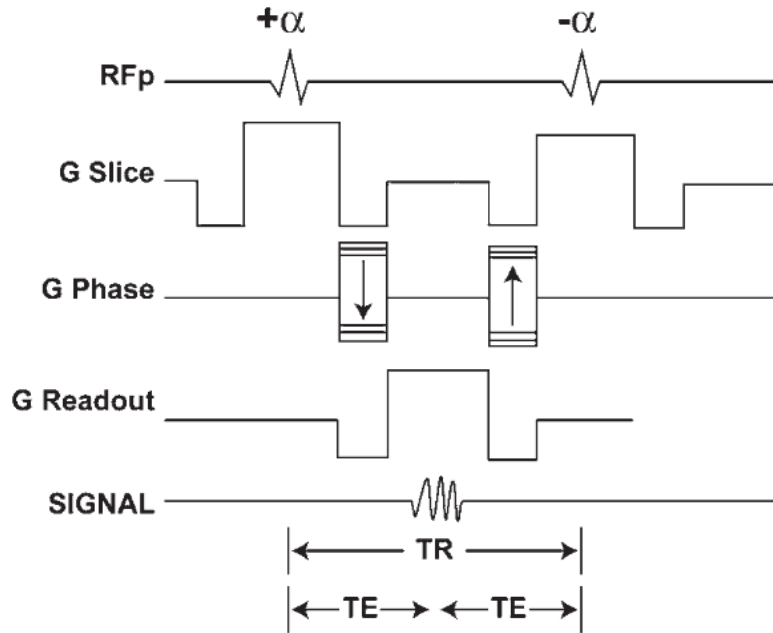


Figura 2.5: Diagrama de pulsos de uma sequência EG b-SSFP. Os gradientes, ao longo dos três eixos, estão perfeitamente equilibrados de forma que, entre dois pulsos de RF, a soma das áreas do gradiente no sentido positivo igualam as áreas negativas. Fonte da imagem: Chavhan *et al.* (2008).

Com as sequências *Steady-State*, a S/R intrínseca é bastante elevada, maior que nas EG *-Spoiled*, permitindo aquisições mais rápidas, através do aumento da BW e diminuição do TR (Elster e Burdette, 2001; Bernstein *et al.*, 2004; Brix *et al.*, 2008; Chavhan *et al.*, 2008; Lin, 2012).

No que diz respeito ao grupo de sequências b-SSFP, uma das que tem tido uma grande implementação na prática clínica, pela sua rapidez e características únicas de sinal, é a **sequência TrueFISP**.²⁴

O contraste na sequência TrueFISP, e nas sequências b-SSFP, em geral, resulta da contribuição de efeitos $T1$, $T2$, do TR e do FA, sendo essencialmente baseado na relação $T2/T1$ do

²⁴Este é o acrónimo utilizado pela Siemens. Na GE é denominada FIESTA (*Fast Imaging Employing Steady-State Acquisition*), na Philips como bFFE (*balanced Fast Field Echo*) e na Toshiba chama-se trueSSFP.

tecido. Nestas sequências, a gordura e o líquido podem apresentar IS elevada (hipersinal), pois têm valores de T1 e T2 muito diferentes mas uma relação T2/T1 semelhante (veja-se Scheffler e Lehnhardt, 2003). De acordo com estes autores, esta sequência é, de todas, a que apresenta a mais elevada S/R por unidade de tempo. Esta será mesmo a grande vantagem da sequência trueFISP que pode ser maximizada com a utilização de FA elevados (cerca 50° a 80°). A C/R é também superior à observada nas sequências EG-*Spoiled*. Outra vantagem é a sua quase insensibilidade a artefactos de movimento e de fluxo (Scheffler e Lehnhardt, 2003). A sua utilização é frequente no âmbito de estudos cardíacos (avaliação morfológica e funcional), estudos abdominais, estudos angiográficos, em doentes pouco colaborantes e em RM fetal. Os tempos de aquisição curtos permitem efetuar estudos em apneia, importante nos estudos abdominais; a nível cardíaco proporcionam um excelente contraste entre o sangue e o miocárdio; proporciona bom contraste entre gordura e músculo e entre compartimentos com líquido e tecido circundante (Scheffler e Lehnhardt, 2003; Chavhan *et al.*, 2008). Na atualidade, é possível obter imagens em tempo real com esta sequência (com preenchimento radial ou espiral) o que é relevante, por exemplo, no contexto da RM cardíaca (Nayak e Hu, 2005; Nayak *et al.*, 2005).

No entanto, a utilização de FA elevados com TR curto, em campos de elevada intensidade dará origem a um aumento do SAR podendo facilmente ser atingidos os limiares recomendados (Scheffler e Lehnhardt, 2003; Carroll *et al.*, 2006). Os grandes problemas desta sequência são os artefactos designados como *banding artifacts* ou *dark-phase dispersion bands* (bandas lineares com baixo sinal) e a sua sensibilidade a efeitos *off-resonance*, em geral. Todavia, a possibilidade de utilizar TR cada vez mais curto (atualmente possível) e a utilização de técnicas adequadas de *shimming* podem minimizar estes artefactos (Elster e Burdette, 2001; Carroll *et al.*, 2006).

2.1.4.2 Técnicas *Single-Shot* e Imagem Eco Planar

Com o intuito de acelerar ainda mais as aquisições surgiram as técnicas *Single-Shot* (doravante designadas como SS). Nalgumas aplicações clínicas (e.g., ColangioPancreatografia por RM - CPRM, estudos de difusão, perfusão) estas técnicas podem substituir com benefícios óbvios (melhor resolução temporal e menos artefactos devido a movimentos respiratórios) as técnicas *Multi-Shot*.²⁵ No entanto, esta substituição nem sempre é possível uma vez que a qualidade das imagens é, de alguma forma, degradada com estas técnicas (Rybicki e Mulkern, 2001).

Um SS é, por definição, uma técnica que permite adquirir todo o espaço K, necessário para reconstruir uma imagem, após a aplicação de um único pulso de excitação de RF, seguido por uma cadeia de ecos (Rybicki e Mulkern, 2001; Bernestein *et al.*, 2004; Dietrich, 2007a; Brix *et al.*, 2008). De entre as técnicas SS destacam-se as técnicas SS Eco de gradiente (SS-EG) e as SS *Fast Spin Eco* (SS-FSE). De forma simplificada, usualmente associam-se as técnicas SS-EG à técnica EPI. Também o SS-FSE é vulgarmente designada como *Rapid Acquisition with Refocused Echoes* ou *Rapid Acquisition with Relaxation Enhancement* (RARE).

A técnica SS-FSE pode ainda ser combinada com uma matriz *Half-Fourier* (HF), também

²⁵Consiste na aquisição de informação à custa de vários *Shots*, cada um dos quais corresponde a um pulso de excitação.

designada por $1/2$ NEX, permitindo reduzir o ETL e reduzir o tempo de aquisição para metade (aproximadamente). A utilização da matriz HF aproveita o facto de o espaço K apresentar simetria hermitiana (ver figura 2.6). Desta forma, é necessário adquirir pouco mais de metade (60%) das linhas sendo as restantes obtidas por simetria a partir destas. O preço a pagar pela rapidez conseguida é uma diminuição da S/R (Elster e Burdette, 2001; Brix *et al.*, 2008).

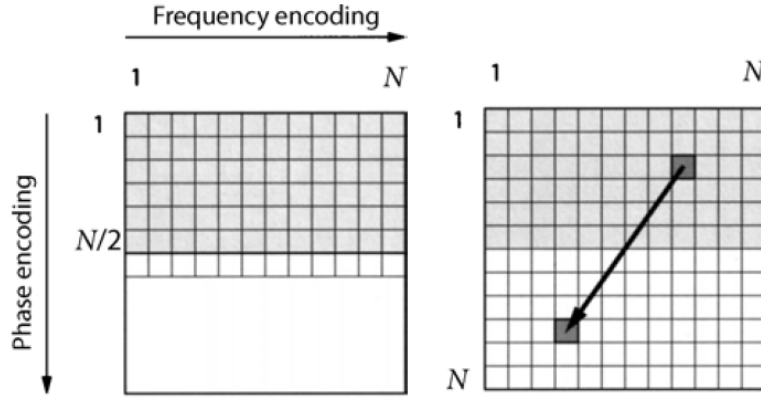


Figura 2.6: Técnica *Half-Fourier*. Com esta técnica, pouco mais de metade dos pontos da matriz de dados é medido (esquerda). Os outros pontos são gerados a partir dos conjugados que foram adquiridos devido à simetria hermitiana do espaço K (direita). Depois disso a imagem é reconstruída através da aplicação da transformada discreta de Fourier 2D à matriz de dados. Fonte da imagem: Brix *et al.* (2008).

A técnica EPI, por vezes chamada sequência Eco-planar, é um método ultra-rápido de aquisição do espaço K. Como já referido anteriormente, apesar de descrita há já alguns anos só mais recentemente foi implementada devido a limitações no desempenho dos gradientes.

Com a técnica EPI, na direção de leitura (direção de codificação da frequência), os gradientes oscilam rapidamente entre valores positivos e negativos permitindo a aquisição de amostras do espaço K ao longo de linhas segundo direções opostas. O preenchimento do espaço K poderá ser em *zig-zag* (utiliza um gradiente de pequena intensidade mas constante na direção de codificação da fase) ou em *blip* (gradiente forte, aplicado durante um curto espaço de tempo - $< 200\mu s$) quando o gradiente de leitura está no zero, na extremidade do eixo K_x . Cada *blip* muda a linha do espaço K que vai ser preenchida como se fosse um degrau. O modo *blipped EPI* é o que é mais frequentemente utilizado, nos dias de hoje, uma vez que obtendo um padrão de amostragem do espaço K ao longo de linhas paralelas, torna mais fácil a aplicação da TF (Elster e Burdette, 2001; Carroll *et al.*, 2006; Brix *et al.*, 2008).

A resolução espacial usualmente conseguida com a técnica SS-EPI ronda os 2 a 3 mm (dimensão do *pixel*) com FOV de 250 mm e matrizes de aquisição de 128×128 , podendo melhorar recorrendo à utilização de técnicas de imagem paralela ou técnicas *Multi-Shot*. Como se ilustra na figura 2.7, o modo EPI pode ser implementado quer com sequências EG (EG-EPI ou FID-EPI) quer com SE (SE-EPI). Para obter um SE-EPI (ver imagem da direita na figura 2.7), é aplicado um esquema (90° - 180° - EPI) permitindo obter contraste T2 muito semelhante ao obtido com uma sequência SE convencional, através da manipulação do TR (Liang e Lauterbur, 2000; Bernstein

et al., 2004; Brix *et al.*, 2008; Hashemi *et al.*, 2012).

O grande problema do modo de aquisição EPI é a sua grande suscetibilidade para artefactos, de natureza variada, tais como: *Nyquist ($N/2$) ghosting*, erros introduzidos pelos gradientes, *off-resonance*, falta de homogeneidade do campo, entre outros. Ainda assim, quando aplicada com um sequência SE é menos sensível a artefactos de suscetibilidade magnética do que com EG-EPI (Brix *et al.*, 2008).

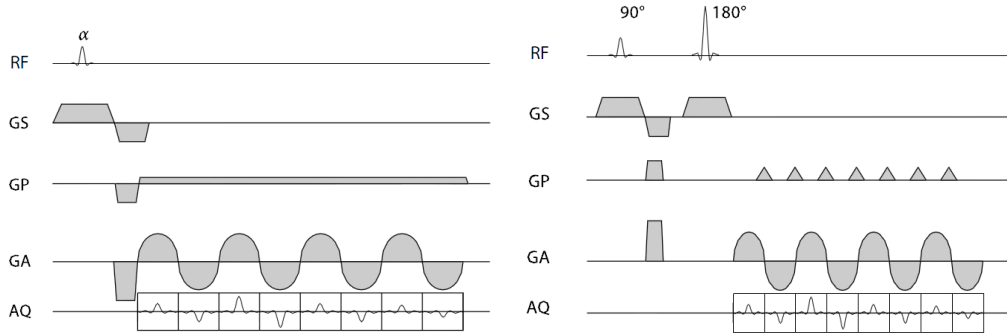


Figura 2.7: Modo de aquisição com esquema de leitura eco-planar. À esquerda EG-EPI e à direita SE-EPI. Fonte das imagens: Brix *et al.* (2008).

A escolha entre as técnicas SS-FSE e SS-EPI depende das aplicações clínicas, uma vez que os artefactos associados a cada uma e os requisitos nos diferentes estudos também o são. A técnica SS-EPI é particularmente utilizada em estudos de perfusão e difusão cerebral, e em técnicas de RMf. Apesar de poder ser também utilizada a nível abdominal ou torácico, os artefactos de suscetibilidade magnética e desvio químico podem revelar-se um grande problema. Com uma resolução temporal mais limitada, as técnicas SS-FSE permitem, nestas estruturas, uma qualidade de imagem superior (Rybicki e Mulkern, 2001; Dietrich, 2007a).

Como veremos de seguida, estas sequências e técnicas de aquisição ultra-rápida podem ainda ser implementadas utilizando estratégias mais eficientes de preenchimento do espaço K (e.g., *SS-Espiral*) ou conjugadas com imagem paralela. A utilização conjunta de técnicas de imagem paralela com técnicas SS, não só permite diminuir o tempo de aquisição como apresenta potencial para melhorar a qualidade de imagem através da minimização de alguns artefactos (Dietrich, 2007a).

2.1.4.3 Trajetórias de preenchimento do Espaço K

Antes de avançar para as diferentes possibilidades de preenchimento do espaço K é pertinente efetuar uma breve introdução relativa à amostragem do espaço K e as implicações para a imagem.

Os dados brutos (*raw data*), ou seja, a informação adquirida antes da reconstrução por inversão da Transformada discreta de Fourier (*Discret Fourier Transform* - DFT) através da aplicação do algoritmo *Fast Fourier Transform* (FFT), pode ser visualizada como uma matriz bidimensional (ou tridimensional nas aquisições 3D) de frequências espaciais (K_x e K_y). Este espaço, no domínio da frequência, denomina-se espaço K.

O centro do espaço K (frequências espaciais mais baixas) é adquirido quando são aplicados gradientes de codificação da fase mais fracos. No entanto, é nessa fase que a intensidade do sinal RM é mais elevada, sendo determinante para o contraste da imagem. Por outro lado, os sinais mais fracos na periferia do espaço K (frequências espaciais mais elevadas) correspondem à aplicação de gradientes de codificação de fase mais intensos e são responsáveis pelo detalhe da imagem (Liang e Lauterbur, 2000; Elster e Burdette, 2001; Wieben, 2007).

Depois de efetuado o preenchimento deste espaço, é necessário aplicar a DFT para que os dados brutos sejam convertidos para o espaço da imagem. A figura 2.8, mostra de que forma o número de amostras, o padrão e a frequência de amostragem influenciam a qualidade da imagem obtida. Uma amostragem limitada às baixas frequências dará lugar a uma imagem com fraca resolução espacial. A subamostragem ao longo de uma determinada direção dá origem a artefactos de *aliasing* nessa direção e uma amostragem não uniforme dará lugar a uma imagem com artefactos aleatórios (*random noise* - imagem da direita) sobre toda a imagem (Lustig *et al.*, 2008).

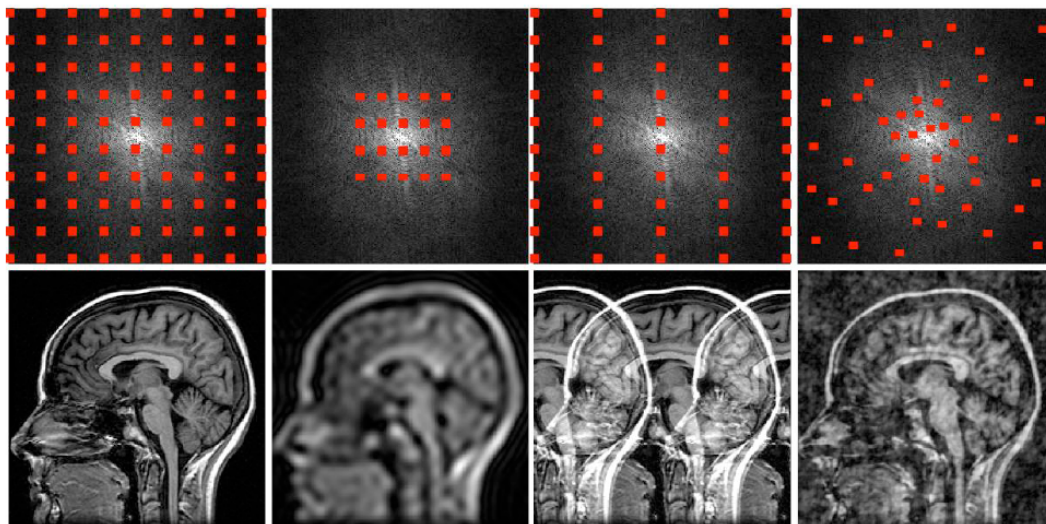


Figura 2.8: Relação entre o padrão de amostragem no domínio da frequência espacial (espaço K) e o resultado no domínio da imagem. A resolução espacial é determinada pela extensão do espaço K (altas frequências) e o FOV pela densidade de amostragem neste espaço. Fonte da imagem: Lustig *et al.* (2008).

A amostragem do espaço K pode ser efetuada com recurso a diferentes trajetórias e padrões de amostragem. A figura 2.9 apresenta diferentes trajetórias e padrões de amostragem do espaço K para aquisições 2D (em cima): cartesiana, cartesiana EPI *blipped*, radial e espiral (da esquerda para a direita). O conceito pode ser estendido a aquisições 3D com a utilização de *stacks* de radiais ou de espirais (em baixo).

Técnicas cartesianas

O preenchimento **cartesiano do espaço K** consiste na aquisição de linhas paralelas do espaço K, podendo ocorrer numa matriz bidimensional ou tridimensional. De acordo com vários autores

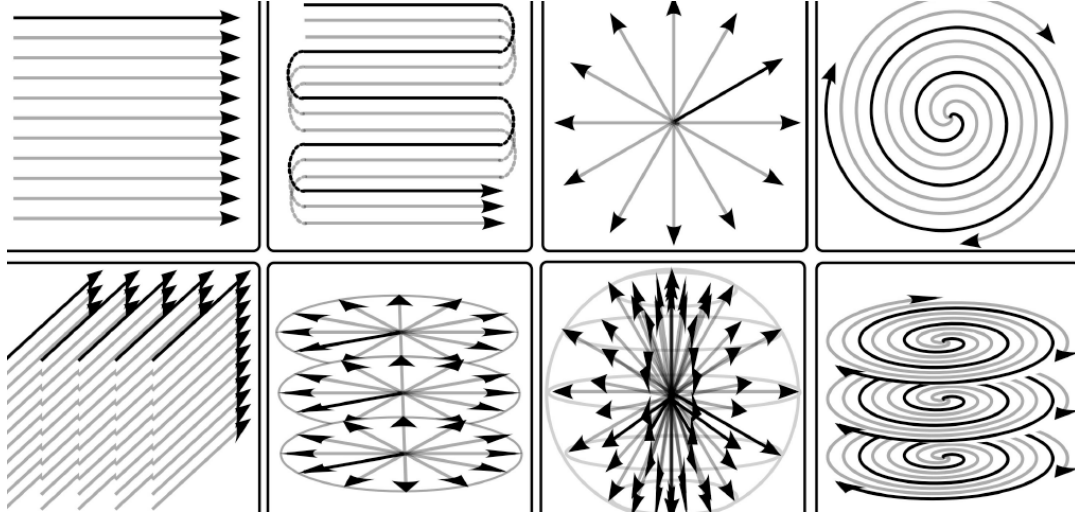


Figura 2.9: Exemplos de trajetórias de preenchimento do espaço K. Em cima aquisições 2D (da esquerda para a direita): cartesiana, cartesiana EPI *blipped*, radial e espiral. Em baixo (esquerda para a direita): cartesiana 3D, range de radiais, radial 3D, range de espirais. Fonte da imagem: Lustig *et al.* (2008).

(Carroll *et al.*, 2006; Lustig *et al.*, 2008; Kim *et al.*, 2010), a trajetória cartesiana é de longe a mais popular de todas e aquela que está implementada na maior parte das sequências que se utilizam, por rotina, no contexto da prática clínica. No entanto, há sequências que, dada a sua natureza, preenchem cartesianamente o espaço K de uma forma temporalmente longa.

Uma das grandes vantagens desta trajetória de preenchimento reside no facto do processo de reconstrução da informação ser extremamente simples, contrariamente às técnicas de preenchimento radial ou espiral, que requerem algoritmos de reconstrução adequados (e.g., *gridding*) ou outros, efetuados na maior parte das vezes *offline*. Na realidade, com a técnica cartesiana basta a aplicação da DFT inversa aos dados brutos para obtenção da imagem, de forma direta e bastante rápida, não implicando nenhuma espécie de interpolação espectral. Para além disso, as reconstruções são robustas às várias fontes de imperfeição inerentes ao sistema (Lustig *et al.*, 2008; Uecker *et al.*, 2010b; Kim *et al.*, 2010).

No entanto, é muito sensível a artefactos de movimento se comparada, por exemplo, com as técnicas de preenchimento radial, o que pode ser um problema quando se trata do acompanhamento de processos dinâmicos em que há movimento do objeto ao longo do tempo.

Técnicas não cartesianas

O recurso a técnicas de **preenchimento radial** do espaço K (reconstrução a partir de projeção)²⁶ tem recebido grande atenção nos últimos anos, principalmente para aquisição de imagem em tempo real, no âmbito de imagem cardíaca, produção de fala, estudos de deglutição, entre outros (Uecker

²⁶Consequência do teorema da secção central tão conhecido em tomografia. Relação entre a transformada de Fourier e transformada de Radon (projeção) numa dada direção e a TF da imagem ao longo dum perfil do espaço K com a mesma direção da projeção.

et al., 2010b; Zhang *et al.*, 2012; Niebergall *et al.*, 2012). Curiosamente, a técnica de reconstrução por projecção foi das primeiras a ser utilizada para reconstruir as imagens de RM. No entanto, devido à baixa homogeneidade do campo magnético e à fraca linearidade dos gradientes que equipavam os *scanners* por volta dos anos 80, a reconstrução com esta técnica dava origem a uma série de artefactos, tendo sido, na altura, abandonada. Na actualidade, com os problemas de homogeneidade de campo e de linearidade dos gradientes ultrapassados, tem recebido novamente atenção, particularmente para algumas aplicações clínicas (Bernstein *et al.*, 2004; Carroll *et al.*, 2006; Uecker *et al.*, 2010a).

Com a trajetória radial, o espaço K é amostrado através de linhas radiais (obtidas a diferentes ângulos) que passam pelo centro do espaço K (radial *spokes*) e não com recurso a uma grelha ou matriz retangular como acontece no preenchimento cartesiano clássico. Desta forma, quer o centro, quer a periferia do espaço K são amostrados a cada TR, minimizando o tempo entre a aquisição da parte central do espaço K e as altas frequência à periferia. Este facto é importante, por exemplo, nos estudos angiográficos, em que a qualidade de imagem é fortemente dependente da aquisição da região central do espaço K, durante a chegada do meio de contraste. A utilização de TE curto, a possibilidade de melhorar a resolução temporal e, principalmente, a reduzida sensibilidade ao movimento, são as vantagens mais apontadas.

Como a região das baixas frequências espaciais apresenta uma densidade de amostragem mais elevada que a das altas frequências, ocorrem artefactos resultantes de amostragem insuficiente nestas frequências espaciais. No entanto, devido à sua incoerência, estes são visualmente menos perceptíveis, prejudicando menos a qualidade da imagem que os originados por subamostragem no preenchimento cartesiano. Como a aquisição de menos *spokes* radiais não contribui para a diminuição da resolução espacial, nem tem implicações a nível do FOV, a aquisição pode ser significativamente subamostrada, permitindo reduzir muito o tempo de aquisição (Uecker *et al.*, 2010b; Zhang *et al.*, 2010a).

O espaço K também pode ser preenchido recorrendo a uma **trajetória espiral**, sendo por vezes designada como **imagem espiral** (*spiral imaging*). De forma semelhante à técnica EPI, este método de preenchimento do espaço K pode ser conjugado com diferentes tipos de sequência (EG, SE ou IR) e é utilizado com o intuito de acelerar o processo de aquisição de imagem (Elster e Burdette, 2001; Bernstein *et al.*, 2004).

Nesta técnica, é utilizado um par de gradientes sinusoidais durante o período de leitura da sequência (Elster e Burdette, 2001; Liang e Lauterbur, 2000). Para originar a espiral é necessário utilizar gradientes com elevado desempenho, particularmente no que diz respeito à sua taxa de variação, que deve ser elevada. Ainda assim, não requer um desempenho dos gradientes tão exigente como a técnica EPI. O número de *Shots* ou *interleaves* necessários numa aquisição, não é, em geral, muito elevado, fazendo com que esta seja uma das trajetórias mais eficientes de preenchimento do espaço K (Bernstein *et al.*, 2004; Lustig *et al.*, 2008; Kim, 2010).

A trajetória inicia-se no centro do espaço K e progride, em forma de espiral, até à periferia, fazendo com que as baixas frequências (contraste) sejam adquiridas no início do período de leitura quando a perda de sinal devido a efeitos T2* é ainda pequena. Por esta razão, os erros

de fase devido a movimento são minimizados. Assim, a sua utilização pode ter grande impacto quando se pretenda adquirir de forma rápida, com menos artefactos devido a fluxo ou movimento, como por exemplo, nos estudos de RM funcional, nos estudos de perfusão e de fluxo (Elster e Burdette, 2001), nos estudos cardíacos ou mesmo de produção de fala/deglutição (Kim, 2010). As trajetórias em espiral são afetadas, mais do que por artefactos devido a erros de fase, por *blurring* com consequente perda de resolução. Este *blurring* é devido a efeitos *off-resonance*, que se propagam por toda a imagem (forma cilíndrica) e não a uma frequência bem localizada.

Outra das condicionantes da técnica espiral reside no facto de a informação obtida a partir destas trajetórias não poder ser diretamente utilizada, sendo necessário recorrer a algoritmos, matematicamente complexos, de reconstrução (*gridding reconstruction*), antes que possa ser aplicada a transformada de Fourier (Liang e Lauterbur, 2000; Elster e Burdette, 2001; Carroll *et al.*, 2006; Kim, 2010). No início da sua aplicação, o tempo despendido na reconstrução das imagens condicionava a sua utilização, por exemplo, quando se pretendia obter imagem em tempo real (Bernstein *et al.*, 2004). No entanto, esse problema foi ultrapassado e têm vindo a ser reportados estudos de RM em tempo real (com elevado número de *frames/s*), nomeadamente na área da produção de fala, que recorrem a esta trajetória (Narayanan *et al.*, 2004; Byrd *et al.*, 2009; Kim *et al.*, 2010, 2011).

2.1.4.4 Técnicas de subamostragem e outras técnicas para acelerar a aquisição

Estão descritas na literatura várias técnicas de subamostragem do espaço K, sendo umas mais comuns que outras. Uma destas técnicas, que já foi atrás referida, é a técnica HF. As técnicas de imagem paralela também podem ser consideradas como um caso particular de subamostragem do espaço K. Outras técnicas muito utilizadas, particularmente em estudos cardíacos ou quando se pretende obter imagens em tempo real, são as técnicas de *View Sharing*. Como se pode deduzir do próprio nome, a informação adquirida é partilhada por diversas *frames* ao longo do tempo, permitindo atingir *frame rate* que dificilmente seria conseguida de forma direta. As técnicas de *View Sharing* são geralmente utilizadas com segmentação do espaço K, podendo ser conjugadas com trajetórias cartesianas e não cartesianas (Wieben, 2007).

Uma das mais populares técnicas de *View Sharing* é a reconstrução por *sliding-windows*. Numa reconstrução clássica, uma imagem pode ser reconstruída sempre que todos os segmentos tenham sido adquiridos. Em alternativa, este processo pode ser efetuado de forma intermitente, obtendo os segmentos em falta, a partir das imagens adjacentes que foram completamente amostradas utilizando *sliding windows*. No entanto, se houver movimento entre as aquisições, o que será normal na monitorização de processos dinâmicos, existirão inconsistências na informação do espaço K, o que dará origem a artefactos de reconstrução (Wieben, 2007; Kim, 2010).

Apesar desta técnica permitir reconstruções com uma *frame rate* elevada, não aumenta a resolução temporal verdadeira, uma vez que esta é determinada pelo tempo que demora a adquirir uma imagem completa (Bernstein *et al.*, 2004; Wieben, 2007; Kim, 2010). A figura 2.10 ilustra o processo de reconstrução com *sliding windows* numa trajetória linear. A utilização de trajetória espiral ou radial com as *sliding windows* torna-se ainda mais vantajosa, uma vez que as

linhas centrais do espaço K, responsáveis pelo contraste na imagem, são amostradas com maior frequência.

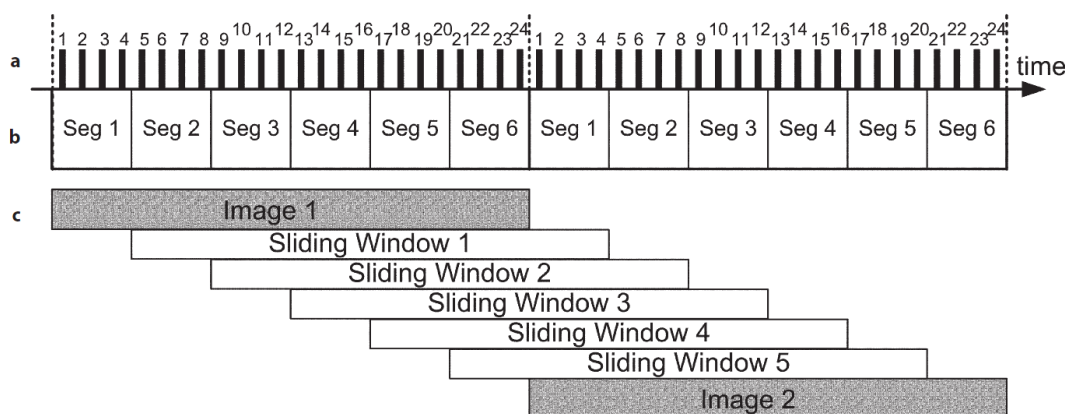


Figura 2.10: Técnica de *View Sharing* com reconstrução por *sliding windows*. Os passos de codificação da fase são ordenados sequencialmente no espaço de *Fourier* e agrupados em seis segmentos (a). Os segmentos são adquiridos repetidamente (b), as imagens podem ser reconstruídas a cada seis segmentos (c), ou geradas de forma intermitente utilizando reconstrução com *sliding windows*. Fonte da imagem: Wieben (2007).

Para além das técnicas referidas, há outras que permitem obtenção e reconstrução de imagem de forma extremamente rápida, tais como *Keyhole imaging*, *Zero-filling* ou *Zero-padding*, etc (Elster e Burdette, 2001; Carroll *et al.*, 2006; Wieben, 2007).

2.1.4.5 Imagem paralela

Como referimos anteriormente, a introdução das antenas *phased-array* constituiu um avanço muito importante na tecnologia de captação de sinal, contribuindo para melhorar a S/R. Esta tecnologia veio, mais tarde, viabilizar a implementação das técnicas de aquisição paralela - *Parallel imaging* (PI) o que veio permitir, desde logo, uma redução muito significativa nos tempos de aquisição com um impacto muito grande em variadíssimas aplicações clínicas: técnicas que exigem aquisições em apneia (e.g., vários estudos abdominais), estudos angiográficos e sempre que seja necessário monitorizar de forma dinâmica os processos fisiológicos.

Pelo facto de este conceito e tecnologia estar, nos dias de hoje, perfeitamente integrado na prática clínica (Blaimer *et al.*, 2004), dedicar-lhe-emos um pouco mais de destaque, providenciando uma visão global do que é a imagem paralela, considerando os pontos comuns e as diferenças entre as várias técnicas (já implementadas) e apontando as vantagens e limitações associadas à sua utilização.

Desde a apresentação do conceito teórico da imagem paralela, que surgiu por volta dos anos noventa, até à sua implementação na prática com as primeiras propostas - SMASH proposta por Sodickson e Manning (1997) e SENSE²⁷ apresentada por Pruessmann *et al.* (1999) - decorreram

²⁷Para não sobrecarregar o texto optou-se por utilizar apenas os acrónimos. Em todo o caso apresentamos o nome, por extenso, das várias técnicas: SENSE-SENSitivity Encoding; m-SENSE - modified SENSE; SMASH -

alguns anos. A partir destas foram desenvolvidas outras técnicas, algumas das quais constituem evoluções das primeiras propostas (e.g., técnica GRAPPA apresentada por Griswold *et al.*, 2002), onde se tentaram ultrapassar problemas de reconstrução da imagem, com o intuito de minimizar artefactos.

De forma genérica, a imagem paralela tira partido da informação espacial contida nos perfis de sensibilidade de uma antena com múltiplos recetores, substituindo, parcialmente, a demorada tarefa de codificação espacial da informação efetuada pela ação dos gradientes de campo magnético. Desta forma, o número de passos de codificação de fase necessários para o preenchimento da matriz pode ser consideravelmente reduzido, acelerando a aquisição, mas mantendo a resolução espacial e o contraste da imagem (Blaimer *et al.*, 2004; Sodickson, 2006). De acordo com uma revisão recente da literatura (Deshmane *et al.*, 2012), as diferentes técnicas de PI partilham algumas características que passamos a enumerar:

1. o espaço K é subamostrado na direção de codificação de fase e na direção de codificação ao longo da partição (nas aquisições 3D), como forma de reduzir o tempo de aquisição. O fator de aceleração ou redução (R) corresponde à redução no tempo de aquisição;
2. os dados são adquiridos recorrendo à informação obtida por um conjunto de bobines recetoras independentes;
3. é necessário recorrer a algoritmos específicos que utilizam a informação dos perfis de sensibilidade de várias antenas para reconstruir a imagem, idealmente sem artefactos, a partir dos dados subamostrados.

Apesar da variedade de técnicas, é consensual entre os autores (Heidemann *et al.*, 2003; Blaimer *et al.*, 2004; Bernestein *et al.*, 2004; Pruessmann, 2006; Deshmane *et al.*, 2012) que este vasto leque de técnicas pode ser agrupado em dois grandes grupos²⁸, tendo em atenção a informação utilizada para as várias etapas de reconstrução: um grupo que engloba as técnicas que operam no domínio da imagem (e.g., SENSE, PILS) e um segundo grupo que contempla as técnicas que operam no domínio da frequência ou do espaço K (e.g., SMASH e GRAPPA).

Nas técnicas que operam no **domínio da imagem**, é necessário, em primeiro lugar, obter os perfis de sensibilidade das antenas (calibração). Depois são obtidas imagens parciais a partir da informação adquirida por cada elemento da antena. Estas imagens parciais, obtidas com FOV retangular apresentam artefactos de *aliasing*. A reconstrução do FOV completo é efetuada pela combinação das imagens parciais, utilizando os perfis de sensibilidade respetivos e eliminando os artefactos de *aliasing*. A aquisição pode ser acelerada por um fator igual ao do número máximo de elementos da antena.

Nos métodos de reconstrução que operam no **domínio do espaço K**, não é necessário obter os perfis de sensibilidade das antenas. Neste caso, são adquiridas linhas extra (de referência)

Simultaneous Acquisition of Spacial Harmonics; GRAPPA - GeneRalized Auto-calibrating Partially Parallel Acquisition; PILS - Partially parallel Imaging with Localized Sensitivities; SPACE-RIP - Sensitivity Profiles from an Array of Coils for Encoding and Reconstruction In Parallel.

²⁸Poder-se-á considerar um terceiro grupo de técnicas designadas como híbridas (SPACE-RIP e generalized SENSE).

na zona central do espaço K (*autocalibration signal lines* ou *ACS lines*) que funcionam como processo de autocalibração. A aceleração é conseguida pelo facto de alguns passos de codificação de fase serem omitidos. As linhas extra vão constituir o modelo para o cálculo dos coeficientes, a aplicar aos sinais adquiridos, para reconstruir as restantes linhas do espaço K.

A figura 2.11 apresenta a forma como é efetuada a reconstrução da imagem com a técnica GRAPPA: (a) informação obtida por cada uma das antenas (espaço K subamostrado); (b) aquisição de linhas extra de referência (ACS), no centro do espaço K; (c) cálculo dos coeficientes; (d) estes são usados para reconstrução da informação em falta no espaço K e, após aplicação da TF, é obtida a imagem (e).

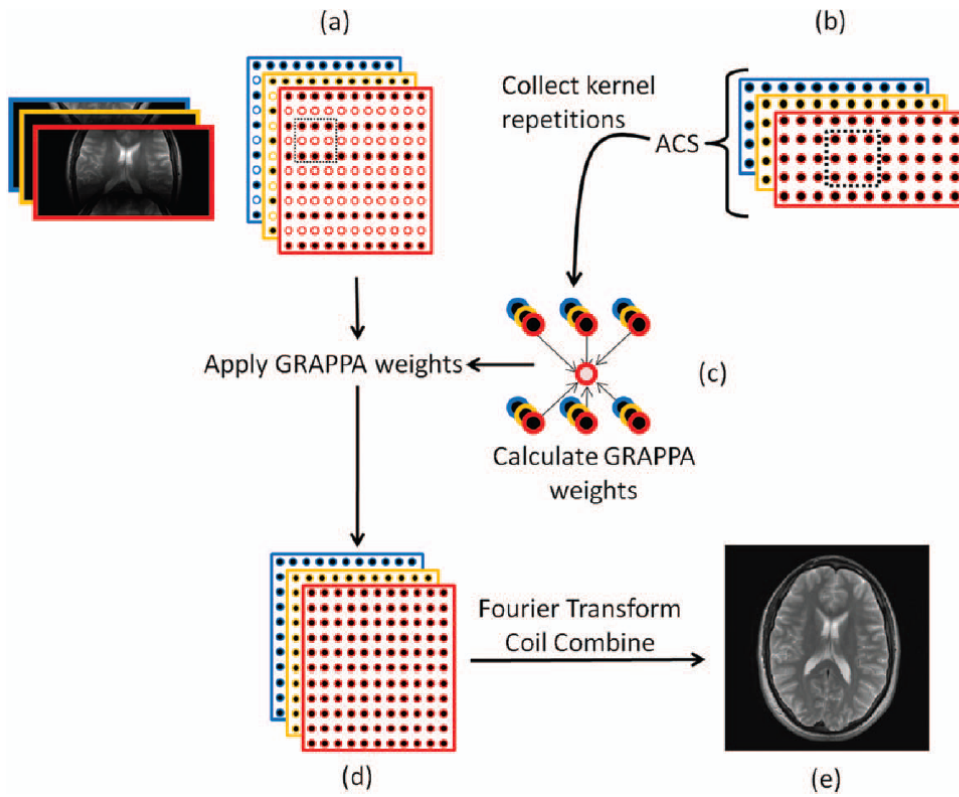


Figura 2.11: Reconstrução da imagem com técnica GRAPPA. Fonte da imagem: Deshmane *et al.* (2012).

A vantagem mais óbvia da imagem paralela é a redução conseguida no tempo de aquisição, uma vez que pode ser diminuído o número de passos de codificação da fase necessários para reconstruir uma imagem. Essa codificação em fase é efetuada, quando não se utiliza imagem paralela, em exclusivo pelos gradientes, através de oscilações mais ou menos rápidas em função dos requisitos da sequência de pulsos. Uma vez que parte desta codificação pode ser efetuada a partir da informação espacial coletada pelas antenas, consegue-se maior rapidez, sem serem necessárias variações tão bruscas dos gradientes. Isto tem um impacto positivo, limitando o surgimento de ENP. O ganho de rapidez também é conseguido sem aumento, ou até com diminuição, do SAR. Muitos autores apontam este facto como outra das mais-valias da imagem paralela, com benefício óbvio nos campos de elevada intensidade onde os elevados valores de SAR constituem um fator

condicionante e uma preocupação de segurança (Dietrich, 2007b; Deshmane *et al.*, 2012) .

Se a rapidez de aquisição não for um parâmetro crítico numa determinada aquisição, o recurso à imagem paralela pode reverter, mantendo o tempo de aquisição, em melhorias na resolução espacial. O estabelecimento de um compromisso entre estes dois parâmetros é a situação mais comum na prática (Blaimer *et al.*, 2004).

Para além destas vantagens, cedo se compreendeu que a imagem paralela poderia contribuir para acelerar ainda mais e/ou melhorar a qualidade de imagem, quando se utilizam técnicas ultra-rápidas *SS e Multi-Shot*, uma vez que pode levar à diminuição dos artefactos de suscetibilidade, *blurring* e distorção geométrica características destas aquisições (Dietrich, 2007b; Blaimer *et al.*, 2004).

No que diz respeito às limitações, para além das condicionantes de ordem técnica (e.g., número de canais mínimo e a sua orientação espacial em função da direção de codificação da fase), o principal problema das técnicas de imagem paralela está associado a um aumento de ruído e à presença de alguns artefactos (*residual aliasing*)²⁹, refletindo-se na qualidade das imagens obtidas. Infelizmente, o preço a pagar por uma maior rapidez é a diminuição da S/R (quanto maior o fator R, menor a S/R), o que pode mesmo inviabilizar ou, em determinadas situações, condicionar um aproveitamento completo das potencialidades destas técnicas. Para todas as situações em que a S/R seja um fator crítico (e.g., baixos campos ou técnicas de alta resolução), a utilização de imagem paralela pode degradar a imagem a ponto de esta não ter qualidade diagnóstica (Dietrich, 2007c; Deshmane *et al.*, 2012). A equação 2.2 mostra de que forma o fator de aceleração afeta a S/R quando se utilizam técnicas PI (forma de cálculo da S/R para a técnica SENSE) (Pruessmann *et al.*, 1999; Deshmane *et al.*, 2012).

$$SNR_{ppi} = SNR_{full} / g\sqrt{R} \quad (2.2)$$

O fator de aceleração pode variar entre 2 e 16, dificilmente se consegue ir para além de fatores acima de 4-5 (Clarke, 2007)³⁰. A figura 2.12 apresenta um exemplo, com utilização de técnica GRAPPA, em que se pode observar a degradação da qualidade da imagem (ruído e artefactos) com o aumento do fator de aceleração.

O surgimento de novos métodos (e.g., *Controlled Aliasing In Parallel Imaging Results IN Higher Acceleration* - CAIPIRINHA) tem contribuído para uma diminuição do impacto dos artefactos de reconstrução nas imagens adquiridas com imagem paralela (Breuer *et al.*, 2006). O conceito foi proposto para melhorar a qualidade das aquisições em 2D, mas é extensível a aquisições 3D através de aceleração, em simultâneo, em duas direções espaciais, ao longo da

²⁹No caso das técnicas SENSE os artefactos residuais de *aliasing* resultam de um perfil de sensibilidade das antenas pouco preciso e no caso das técnicas GRAPPA resultam da utilização de fatores de aceleração elevados (Deshmane *et al.*, 2012).

³⁰Para além deste decréscimo na S/R devido ao fator de aceleração, há ainda a considerar o efeito de amplificação localizada do ruído que é quantificável através do *factor-g*. Esta amplificação do ruído, manifesta-se em forma de banda na parte central da imagem e depende de vários parâmetros como a geometria e número de elementos da bobine, a orientação da imagem, e o fator R. Fatores de aceleração muito elevados conduzem ao aumento do *factor-g*, que idealmente deve ser baixo (Clarke, 2007). O impacto deste fator e a sua otimização na imagem paralela é complexo e está fora do âmbito deste trabalho.

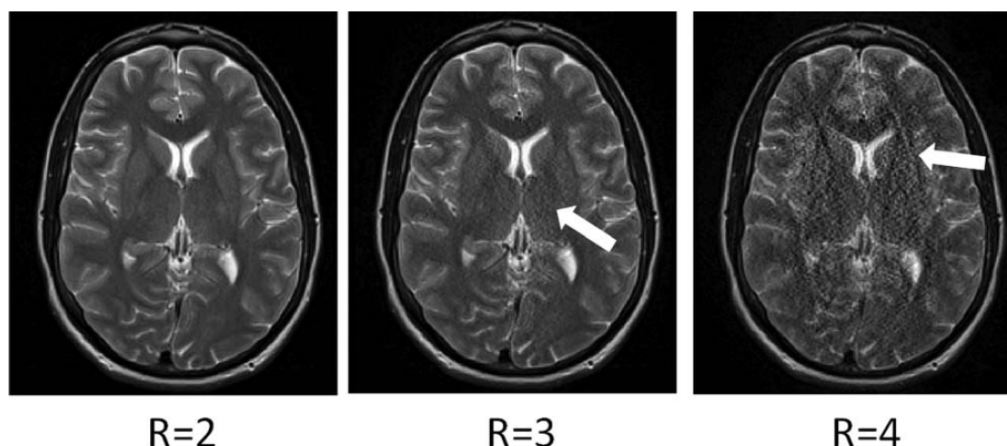


Figura 2.12: Exemplo de imagens axiais do crânio adquiridas com recurso à técnica GRAPPA com diferentes fatores de aceleração ($R = 2$, $R = 3$ e $R = 4$) utilizando 24 linhas de referência (ACS). Total de 12 bobinas recetoras utilizadas para adquirir a informação. Com o aumento do fator de aceleração há um aumento do ruído e surgem artefactos de *aliasing* residuais ($R = 4$). Fonte da imagem: Deshmene *et al.* (2012).

direção de codificação da fase e ao longo da direção da partição (Breuer *et al.*, 2006; Deshmene *et al.*, 2012).

Outro método, recentemente apresentado, com o objetivo de minimizar os artefactos de *aliasing* associados com a técnica GRAPPA convencional, é designado como *Cross-Sampled GRAPPA* (CS-GRAPPA). Os autores propõem que as linhas de referência usadas para a calibração (ACS) sejam adquiridas numa direção perpendicular à das linhas subamostradas. O método melhora o processo de calibração quando são utilizadas poucas linhas de referência (Wang *et al.*, 2012).

Outras propostas, relativamente recentes, incluem a conjugação de várias técnicas como, por exemplo, PI com técnicas de *Compressed Sensing* (Bilgin *et al.*, 2008), a utilização de PI para acelerar aquisições com trajetória em espiral (Heberlein e Hu, 2006) ou radial (e.g., radial-GRAPPA em Seiberlich *et al.*, 2011).

2.1.4.6 Técnicas emergentes

Apesar de já se conseguir adquirir imagens de RM de forma muito rápida, é consensual que algumas aplicações clínicas viriam a beneficiar de uma melhoria na resolução temporal (e/ou espacial). Em geral, são apontadas as situações em que seja necessário acompanhar ou monitorizar processos dinâmicos como acontece na imagem cardíaca, estudos funcionais, angiográficos, fluxo, procedimentos de intervenção, entre outros (Tsao *et al.*, 2003; Huang *et al.*, 2005; Lustig *et al.*, 2008). Há essencialmente duas formas de diminuir o tempo de aquisição e, consequentemente, melhorar a resolução temporal: 1) amostrar de forma rápida e eficiente o espaço K ou 2) adquirir menos informação. A implementação de sequências de pulso ultra-rápidas a par com os avanços na tecnologia

dos gradientes e a exploração de várias formas de codificação da informação contribuíram, no seu conjunto, para adquirir informação de forma mais rápida. Esta foi uma das primeiras abordagens para diminuir os tempos de aquisição. No entanto, por este caminho dificilmente se conseguirão maiores ganhos de rapidez, uma vez que há condicionantes a nível tecnológico (e.g., amplitude e taxa de variação dos gradientes) e os limiares fisiológicos (e.g., ENP) começam a ser atingidos (Tsao *et al.*, 2003; Lustig *et al.*, 2008; Larkman, 2009).

A outra via para obter maior rapidez de aquisição tem sido conseguida às custas de uma redução da quantidade de dados adquiridos, ou seja, através do recurso a várias técnicas de subamostragem do espaço K (e.g., *Half-Fourier* e técnicas de imagem paralela). A subamostragem do espaço K, se existir violação do teorema de *Nyquist*, dará sempre origem a artefactos na imagem reconstruída que prejudicam a sua qualidade. Os fatores de aceleração conseguidos com a utilização de imagem paralela são condicionados pelo número de canais disponíveis e, como já referido, pela S/R, pelo que não é possível acelerar a aquisição tanto quanto seria desejável. A aquisição parcial do espaço K e as técnicas de imagem paralela, de forma um pouco diferente, exploram sempre correlações e/ou simetrias a nível do espaço K. Por outro lado, as técnicas como *Keyhole*, *view sharing* (e.g., *Sliding Windows*) ou UNFOLD exploram apenas correlações temporais para reconstruir as imagens, utilizando várias formas de interpolação. Independentemente da técnica, cada ponto do espaço K é reconstruído sem ter em conta a informação contida em todos os outros pontos (veja-se Tsao *et al.*, 2003).

Uma vez que as imagens dinâmicas referentes a objetos naturais são significativamente correlacionáveis no espaço e no tempo, se forem utilizadas técnicas que exploram, em simultâneo, correlações no espaço e no tempo, será possível melhorar consideravelmente a resolução temporal ou, em alternativa, a resolução espacial (Tsao *et al.*, 2003). Técnicas como k-t BLAST (*Broad-use Linear Acquisition Speed-up Technique*), k-t SENSE (Tsao *et al.*, 2003) e k-t GRAPPA (Huang *et al.*, 2005), que conjugam correlações temporo-espaciais e imagem paralela, têm sido testadas no âmbito de estudos cardíacos, revelando-se muito promissoras. As aplicações podem ser estendidas a estudos abdominais e pulmonares, permitindo acompanhar os movimentos (quasi-periódicos) dessas estruturas, com elevada resolução temporal (Tsao *et al.*, 2003).

Também as técnicas designadas genericamente como *Compressed Sensing-CS* (Lustig *et al.*, 2007, 2008; Kim, 2010) começam a ser aplicadas na área da RM, visando acelerar as aquisições à custa da aquisição de menos informação. As técnicas CS têm implícitos os conceitos de compressão, esparsidade e redundância presentes nas imagens. A sua aplicação implica que algumas condições tenham que ser verificadas, nomeadamente: a) a imagem tem que ter uma representação esparsa num domínio de transformação conhecido (e.g., a imagem tem que ser compressível), b) os artefactos de *aliasing* resultantes de subamostragem do espaço K têm que ser incoerentes (tipo ruído) nesse domínio de transformação e c) tem que ser efetuada uma reconstrução não linear para forçar a esparsidade da imagem representada e a consistência dos dados adquiridos. De acordo com Lustig *et al.* (2008), as características das imagens de RM, particularmente mas não só, relativas a imagens dinâmicas, cumprem estes requisitos. São apontadas algumas áreas em que a utilização de CS é potencialmente aplicável: angiografia 3D, estudos cardíacos, estudos cerebrais.

No entanto, os autores referem que a utilização de técnicas CS com RM, em contexto clínico, está numa fase muito inicial e, apesar de promissoras, estão por resolver/melhorar aspetos cruciais como: 1) a otimização das trajetórias de amostragem, 2) o desenvolvimento de transformações esparsas que conduzam a artefactos espacialmente incoerentes e, em consequência, visualmente inócuos ou irrelevantes, 3) a avaliação da qualidade das reconstruções e sua significância a nível clínico e 4) o aumento da rapidez dos algoritmos de reconstrução (Lustig *et al.*, 2007, 2008).

Nestes últimos anos, a quantidade de técnicas que têm sido propostas, baseadas nestes conceitos, superando alguns dos problemas iniciais, é considerável. As áreas de aplicação que têm merecido mais interesse são os estudos cardíacos e angiográficos. No entanto, a maior parte delas não está ainda implementada na prática clínica diária. Técnicas como k-t FOCUSS (Jung *et al.*, 2009) ou k-t SPARSE-SENSE (Kim *et al.*, 2012; Feng *et al.*, 2012) constituem apenas alguns exemplos destas evoluções. Uma revisão da literatura recentemente apresentada (veja-se Tsao e Kozerke, 2012) proporciona uma visão do estado da arte, no que diz respeito às técnicas de aceleração em RM.

2.2 RM no âmbito de estudos de produção de fala

À semelhança do que aconteceu com outras modalidades de imagem médica - tome-se como exemplo a utilização dos RX na área da produção de fala, que ocorreu escassos anos após a sua descoberta por Röntgen³¹ - também o recurso à RM neste domínio aconteceu não muito tempo depois da introdução da RM na prática clínica. Desde essa altura, os estudos conduzidos nesta área, acompanhando sempre de perto as evoluções desta modalidade imagiológica, têm angariado um grande manancial de informação para várias línguas, para vários sons e com objetivos muito diversificados. Neste sentido, nos parágrafos subsequentes, providenciaremos uma visão global do que tem sido a aplicação desta modalidade imagiológica nos estudos de produção de fala, centralizando-nos um pouco nos aspetos técnicos referenciados em cada um dos estudos e estabelecendo paralelismo com os desenvolvimentos apontadas nas secções anteriores.

As primeiras utilizações da RM na área da produção de fala reportam-se ao final dos anos oitenta e foram conduzidas por Rokkaku *et al.* (1986) e por Baer *et al.* (1987). No entanto, o estudo de Baer *et al.* de 1991 é o que mais frequentemente é referenciado na literatura, pois, pela primeira vez, foi reportada informação 3D relativa ao trato vocal a partir de imagens de RM. Desde essa altura até ao presente, uma grande variedade de técnicas de RM têm sido exploradas neste domínio, desde as mais simples às mais complexas e exigentes (e.g., *tagged MRI* ou RM em tempo real).

A forma como têm sido “catalogados” e divididos estes estudos de RM é bastante diversa. Por exemplo, em Masaki *et al.* (2008), os autores consideram uma divisão em três fases: uma primeira fase de início da utilização da RM na área da produção de fala, que se desenvolve pelos

³¹A descoberta dos Raios X, por William Conrad Röntgen, foi anunciada à comunidade científica em 8 de Novembro de 1895. De acordo com Honda (2002) a sua primeira utilização na área da produção de fala terá sido efetuada por Moeller e Fischer em 1904 com o intuito de observar a posição das cartilagens laríngeas com diferentes níveis de *Pitch*, sendo seguida por Meyer em 1910 para estudar vogais.

anos oitenta e noventa, e inclui apenas articulações passíveis de serem sustidas, devido aos tempos de aquisição muito longos; uma segunda fase com aplicação de técnicas RM-Cine, *tagged MRI* e *Synchronized Sampling Method - SSM* (depois de 1990 e até 2001) e, finalmente, uma terceira fase, que consideram a fase de refinamento das técnicas de aquisição e de visualização (período de 2003-2006), onde se incluem também as técnicas para co-registo das arcadas dentárias, gravação de som com tecnologia adequada e visualização 3D em modo *cine*. Nesta revisão, não é efetuada qualquer referência à possibilidade de adquirir imagens com RM em tempo-real, apesar de existirem na altura variadas publicações que o referem (e.g., Demolin *et al.*, 1997 ou Narayanan *et al.*, 2004).

Outros autores, como Narayanan *et al.* (2004) ou Engwall (2002b, 2003) consideram estudos estáticos, por um lado, e estudos dinâmicos, por outro, tendo em atenção a informação que pode ser retirada das imagens. Narayanan *et al.* (2004) esclarece que a informação dinâmica (relativa ao movimento dos articuladores no tempo) pode ser obtida através de abordagens/técnicas diferentes: 1) um método que permite reconstruir a informação a partir de um número elevado de repetições dos sons (a que chama RM dinâmica) e 2) outra técnica que permite adquirir imagens de forma contínua, em tempo real (ou quase em tempo real).

Apesar de poder existir uma associação cronológica entre as diferentes abordagens e as evoluções técnicas na área da RM, a verdade é que, no presente, todas estas abordagens coexistem, havendo, no entanto, grandes melhorias e aperfeiçoamentos dentro de cada uma delas. Por exemplo, no que diz respeito aos estudos estáticos, o tempo de aquisição que se consegue atualmente nada tem a ver com os minutos reportados nos estudos iniciais para adquirir uma pequena quantidade de cortes. Para além de se adquirir mais informação em menos tempo, a qualidade das imagens é muito superior, abrindo também a possibilidade de novas aplicações no domínio do processamento de imagem, o que não é menos relevante.

Nos estudos em que se pretende obter informação 3D, a quantidade de informação a adquirir é significativa e os tempos de aquisição são mais longos, tornando-se necessário que o som seja sustido por um determinado período de tempo durante a aquisição de imagens, o que não é muito natural e, como será óbvio, não permite obter informação dinâmica. Por esta razão, estes estudos incluem maioritariamente vogais e consoantes (e.g., oclusivas, fricativas, laterais) passíveis de serem sustidas. Por exemplo, sons com características mais dinâmicas que envolvem vibração de articuladores (e.g. vibrantes) não são usualmente incluídos nos estudos estáticos.

Mesmo com limitações, as aquisições estáticas permitem obter informação 2D e 3D de todo o trato vocal, o que é muito relevante, sendo por esse motivo muito utilizadas na área. Na maior parte destes estudos, a informação 3D (e.g., áreas do trato vocal, volumes ou reconstruções 3D de articuladores como a língua ou véu palatino) provém de cortes com orientação axial, efetuados ao longo da faringe, cortes oblíquos na zona do véu palatino e coronais na cavidade oral. Com pequenas diferenças entre estudos, esta foi a abordagem seguida durante alguns anos por diferentes autores (veja-se a tabela 2.2 com exemplos de alguns estudos estáticos), mas encontra-se, também, em investigações mais recentes como, por exemplo, em Rua e Freitas (2007). Na figura 2.13, apresenta-se um exemplo, retirado de Badin *et al.* (2002), de cortes efetuados em vários planos para obter reconstrução 3D da língua.

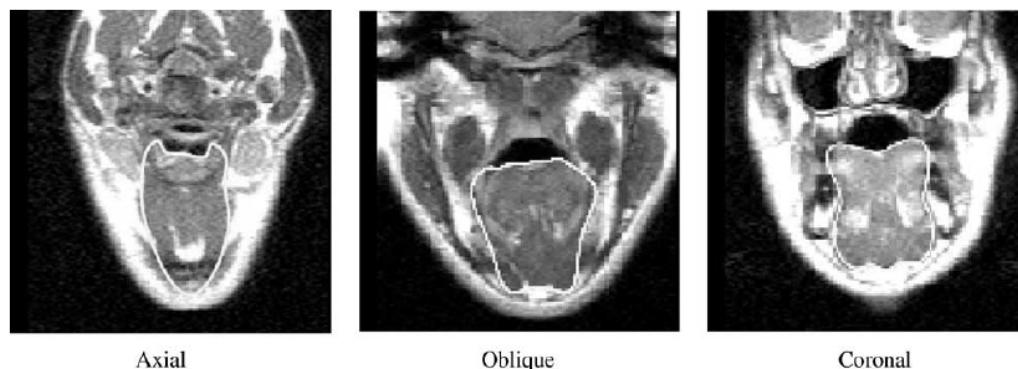


Figura 2.13: Exemplo de aquisição 2D em vários planos (axial, oblíquo e coronal) para obtenção de informação 3D da língua. Fonte da imagem: Badin *et al.* (2002).

O recurso a verdadeiras aquisições 3D³², no âmbito destes estudos, é menos frequente e mais recente. Um dos primeiros estudos que utilizou esta abordagem foi o estudo de Martins (2007) para estudo dos sons do PE, com recurso a uma sequência EG-*Spoiled* ultra-rápida (3D FLASH-VIBE), conseguindo tempo de aquisição de 18 s com informação de todo o trato vocal. O recurso a técnicas de imagem paralela e de outras que exploram a esparsidade do espaço K (e.g., *Compressed Sensing*) começam a ser referidas na área (e.g., Kim *et al.*, 2009). Estes autores conseguiram fatores de aceleração consideráveis com uma aquisição em volume. Shosted *et al.* (2012b) adaptam o protocolo otimizado por Martins (2007) e publicado em Martins *et al.* (2008) para obtenção de dados 3D relativos à cavidade faríngea em falantes do árabe. Recentemente, Ventura *et al.* (2013) utilizam também a sequência FLASH-3D num estudo com o objetivo de avaliar a morfologia do trato vocal, durante o canto, em cantores profissionais.

No que diz respeito a outros aspetos técnicos, têm sido utilizados equipamentos com campos de intensidade diversa, desde 0.15 até 3T. As sequências mais utilizadas foram as SE e TSE nos estudos mais antigos e as EG nos mais recentes. A utilização de bobines de crânio e pescoço, individualmente, ou em conjunto é o mais habitual.

Note-se a extraordinária melhoria nos tempos de aquisição desde o primeiro estudo, em que foram necessários cerca de **30 minutos** para obter uma configuração do trato vocal, a partir de múltiplas repetições, até à atualidade, em que uma configuração do trato pode ser obtida em apenas **7s**, numa única produção, evitando inconsistências aquando da reconstrução da informação.

³²Numa aquisição 3D “verdadeira”, ao invés de se excitar um ou vários cortes, todo um volume de interesse é excitado. Neste modo de aquisição, para além dos dois gradientes que efetuam a codificação da fase e da frequência, no plano, é aplicado um segundo gradiente de codificação da fase, perpendicularmente aos anteriores (ao longo da direção de seleção de corte - eixo dos ZZ), sendo depois aplicada a Transformada de Fourier 3D. A grande vantagem deste modo de aquisição é que o sinal medido integra o contributo de todas as localizações dentro do volume de interesse resultando num incremento da S/R. Desta forma, é possível ter *voxels* de menor dimensão (e.g., matrizes mais elevadas), conseguindo-se melhorar a resolução espacial, com S/R adequada. Posteriormente, permite a realização de reconstruções com espessuras de corte mais finas, em qualquer plano ou orientação. Para além disso, facilita a utilização de outras abordagens de processamento de imagem, nomeadamente, métodos de segmentação verdadeiramente 3D (e.g., 3D *bubbles*). A maior condicionante, principalmente no início da sua utilização, teve a ver com os tempos de aquisição bastante longos. No entanto, nos últimos anos, a implementação de sequências EG e TSE ultra-rápidas, em conjugação com outras técnicas de aquisição rápida (e.g., imagem paralela e/ou *compressed sensing*), abriu a possibilidade de adquirir volumes de dados muito consideráveis com tempos de aquisição significativamente mais curtos.

Tabela 2.2: *Exemplos de estudos estáticos com RM.*

Autores	Informação	B_o	Sequência	TA/cortes	Corpus	Falantes
Baer <i>et al.</i> (1991)	2D/3D	0.15 e 1.5 T	SE	30 <i>min</i> (por som)	vogais	2 (M)
Narayanan <i>et al.</i> (1995) Narayanan <i>et al.</i> (1997b) Narayanan <i>et al.</i> (1997a) Alwan <i>et al.</i> (1997)	2D/3D Cor, Sag, Tra	1.5T	EG-GRASS	1-16 s (cada 4 imagens) 3 mm 40-45 cortes	fricativas vogais laterais róticos	4 (2M/2F)
Demolin <i>et al.</i> (1996) Demolin <i>et al.</i> (2003)	2D/3D Cor, Cor obl, Tra	1.5T	TSE (DP)	14 s (14 cortes.) 13.8 s (18 cortes) 4 mm	10 vogais	4 (2M/2F)
Story <i>et al.</i> (1996)	3D Tra	1.5T	TSE	4 m 16 s (26 cortes)	vogais, oclusivas, cons. nasais	1 (M)
Badin <i>et al.</i> (1998)	2D/3D Cor, Cor obl, tra	1T	SE	43 s (54 cortes) 3.6 mm	consoantes VCV Vogais	1 (M)
Engwall e Badin (1999)	2D/3D Cor, Cor obl, tra	1T	SE	43 s (54 cortes) 3.6 mm	47 conf. trato	1 (M)
Kröger <i>et al.</i> (2000)	3D (Sag) Sag	1T	TSE	21 s (21 cortes) 3.5 mm	vogais longas	1 (M)
Serrurier e Badin (2005b)	3D Sag	1T	–	45 s (25 cortes)	46 conf. trato	1 (M)
Ericsson (2005)	3D Tra, Cor	1.5T	TSE (DP)	13-14 s (15 cortes) 4 mm	9 vogais 2 alofones	2 (1M/1F)
Rua e Freitas (2007)	3D Cor, Sag, Tra	1.5T	TSE-T1	9-10 s (9 cortes) 5-6 mm	vogais consoantes	2 (1M/1F)
Martins (2007)	2D/3D 3DFT (tra)	1.5T	EG-VIBE	18s (64 cortes) 2 mm	vogais consoantes (23 conf. trato)	1 (M)
Kim <i>et al.</i> (2009)	3D 3DFT (sag)	3T	EG	7 s 2 mm	fricativas líquidas	1 (M)

As aquisições dinâmicas abriram a possibilidade de obter informação temporal relativa ao movimento dos articuladores, no entanto há necessidade de efetuar múltiplas repetições do som (tantas quantas o número de passos de codificação de fase), para obter, após reconstrução das imagens, uma pseudo variação temporal do trato vocal. Esse facto constitui, na opinião de alguns autores (Engwall, 2002b; Narayanan *et al.*, 2004), uma limitação desta técnica, uma vez que, mesmo com treino, é difícil para um falante efetuar produções exatamente iguais. O método é semelhante ao modo de aquisição utilizado em RM para estudos cardíacos, através de aquisição sincronizada com o ECG ou com um *trigger* externo. Esta abordagem, com pequenas diferenças, foi seguida em vários estudos (Foldvik *et al.*, 1993; Masaki *et al.*, 1997; Shadle *et al.*, 1999; Mohammad, 1999; Mathiak *et al.*, 2000; Rua e Freitas, 2007).

A possibilidade de obter imagens de RM quase ou efetivamente em tempo real³³, no âmbito dos estudos de produção de fala, revelou-se de extrema importância. Para uma modalidade como a RM, em que o processo de aquisição, codificação e reconstrução de imagem é inerentemente demorado, o desafio foi grande, mas é hoje possível, fruto dos desenvolvimentos que já demos conta nas secções anteriores deste capítulo.

Os primeiros estudos com utilização de RM em tempo real, na área da produção de fala, reportam-se aos anos de 1997 e 2000 e foram conduzidos pela mesma equipa de investigação (Demolin *et al.*, 1997, 2000). Apesar da resolução temporal ser relativamente baixa, estes estudos constituíram uma grande evolução relativamente ao que, até à data, havia sido conseguido na área da produção de fala.

Nos estudos de Demolin *et al.* (1997) e Demolin *et al.* (2000) foi possível obter uma

³³Mais adiante, no capítulo 6 far-se-ão mais algumas considerações relativamente à aquisição de imagem em tempo real.

resolução temporal de 250 e 200 *ms* a que correspondem, respetivamente, 4 e 5 imagens/s. Esta diminuição drástica do tempo de aquisição (para os *ms*) foi conseguida, nesse estudo, à custa da implementação de uma sequência TSE ultra-rápida (TSE-ZOOM)³⁴.

O protocolo de Demolin *et al.* (2000) foi, posteriormente, utilizado nos trabalhos de Delvaux *et al.* (2002a) e Delvaux (2003) para estudo das vogais nasais do francês e por Medeiros e Demolin (2006) para o Português Brasileiro.

Na tabela 2.3, apresentam-se alguns estudos com RM em tempo real realizados neste contexto. No que diz respeito à intensidade do campo magnético nos estudos em tempo real, como se pode observar, esta tem variado entre 1.5 a 3T. A utilização de campos de maior intensidade nos estudos de produção, nos últimos anos, reflete a sua crescente implementação em vários centros e, também, a resolução de alguns problemas relacionados com a qualidade das imagens. Todavia, a utilização de equipamentos de 1.5T, em conjunto com o recurso a antenas dedicadas, trajetórias eficientes, subamostragem dos dados e técnicas de reconstrução adequadas, podem permitir aquisições tão rápidas e com qualidade equiparável às obtidas com campos de 3T. Veja-se, por exemplo, que o estudo mais recente apresentado na tabela 2.3, envolvendo uma verdadeira aquisição volumétrica em tempo real, foi conduzido num campo de 1.5T (Zhu *et al.*, 2013b).

Tabela 2.3: *Exemplos de estudos com RM em tempo real*

Estudo	B_o (T)	Sequência	Trajectória	Frames/s	Esp. Corte (mm)
Demolin <i>et al.</i> (1997)	1.5T	TSE Zoom	Cartesiana	4	8
Demolin <i>et al.</i> (2000)	1.5T	TSE Zoom	Cartesiana	5	8
Mády <i>et al.</i> (2002)	1.5T	EG rápida (FGE)	Cartesiana SENSE	8	8
Engwall (2004)	1.5T	TSE Zoom	Cartesiana	9	6
Narayanan <i>et al.</i> (2004)	1.5T	EG-Spoiled (FSPGR)	Espiral	8/9	3 e 5
Kim <i>et al.</i> (2010)			Slid. windows	24	
Martins (2007)	1.5T	EG bSSFP TrueFisp	Cartesiana	5	8
Echternach <i>et al.</i> (2008)	3.0T	EG bSSFP	Cartesiana	4	11
Echternach <i>et al.</i> (2010)				10	11
Uecker <i>et al.</i> (2010b)	3.0T	EG-Spoiled FLASH	Radial	18 e 30	10
Niebergall <i>et al.</i> (2012)				30	10
Shosted <i>et al.</i> (2012b)	3.0T	EG-Spoiled	Espiral	15	10
Zhu <i>et al.</i> (2013a)	1.5T	EG-Spoiled	Espiral	23.8 (<i>Sl. wind</i>)	3
Zhu <i>et al.</i> (2013b)	1.5T	EG-Spoiled	3D Espiral	8.6	volume

Como se pode constatar, as primeiras sequências a ser utilizadas para obtenção de imagem em tempo real foram as sequências TSE, mas têm sido praticamente substituídas pelas sequências EG ultra-rápidas. Isso pode facilmente ser justificado pela rapidez que estas últimas permitem

³⁴Esta variante da sequência TSE é por vezes designada como ZOOM imaging ou LoLo (*Local Look imaging*) e encontra-se disponível nos equipamentos da Philips. A grande vantagem é que permite utilizar FOV muito pequeno (pelo menos numa das direções) sem que ocorram os inconvenientes artefactos de *aliasing*. Desta forma, pode-se conseguir uma melhoria na resolução espacial ou, em alternativa, aumentar a resolução temporal uma vez que o número de passos de codificação da fase é baixo. Para isto é necessário que a largura de banda do pulso de RF e o gradiente de seleção de corte sejam escolhidos por forma a que a gama de frequências do perfil do corte coincida com o FOV na direção de codificação da fase. A desvantagem é que permite a obtenção de um único corte (Bernstein *et al.*, 2004). Nos estudos em tempo real, na área da produção de fala, geralmente efetua-se apenas um corte sagital, na linha média, pelo que não constitui um grande obstáculo.

e por existir vasto trabalho de otimização com as sequência EG, uma vez que são estas também as mais utilizadas para dar resposta aos exigentes requisitos levantados no âmbito da imagem cardíaca.

Dentro das sequências EG rápidas, tem sido reportada a utilização das sequências b-SSFP com os gradientes perfeitamente equilibrados (TrueFISP na Siemens), sequências Eco de Gradiente (EG)-*Spoiled* (como a sequência FLASH) e com preparação da magnetização como a sequência turbo-FLASH. A imagem eco-planar, apesar de ser das abordagens que permite aquisições mais rápidas, não se tem constituído como uma alternativa viável, devido aos inúmeros artefactos presentes na imagem (Narayanan *et al.*, 2004; Kim, 2010).

A maior parte dos estudos tem recorrido à clássica técnica cartesiana de preenchimento do espaço K. O primeiro estudo que refere a utilização de imagem espiral é o estudo de Narayanan *et al.*, em 2004, conseguindo obter 9 *frames/s* (resolução de 110 *ms*) com espessura de corte de 3 ou 5 *mm*. Com recurso a *sliding windows*, as imagens são reconstruídas, obtendo-se *frame rate* de 24 *frames/s* ou superior. Com pequenas variações nos parâmetros utilizados, esta tem sido a sequência utilizada numa série de estudos posteriores conduzidos pelo grupo da Universidade da Califórnia (SPAN - Speech Production and Articulation kNowledge Group). Será justo destacar o vasto e variado trabalho que tem sido levado a cabo por esta equipa de investigação, na utilização de RM na área da produção de fala, não só no aperfeiçoamento de técnicas relativas à aquisição de imagem, mas também no seu processamento (ver capítulo 3). Muitos destes trabalhos já foram ou serão referenciados ao longo do texto³⁵.

A utilização de técnica de preenchimento radial do espaço K só mais recentemente foi reportada (Uecker *et al.*, 2010b). Este estudo não é especificamente dedicado à produção de fala. Os autores apresentam uma abordagem para acelerar as aquisições de imagem com RM (elevada resolução temporal) que conjuga a utilização de sequência ultra-rápida (FLASH) com uma trajetória radial e subamostragem do espaço K. Foram testados vários protocolos para diferentes aplicações: estudo cardíaco, produção de fala e fluxo turbulento (em fantoma). No que diz respeito às imagens obtidas durante a produção de fala, foram reconstruídas imagens com resolução temporal de 55 *ms* (18 *frames/s*) e 33 *ms* (30 *frames/s*). A espessura de corte foi de 10 *mm*, matriz quadrada de 128 elementos com resolução espacial de 1.5×1.5 *mm*². Para reconstrução das imagens foi utilizado um método de inversão não linear, efetuado *offline*, e filtros de reconstrução, para minimizar artefactos. Posteriormente, com algumas alterações, o protocolo foi aplicado para avaliação da deglutição (Zhang *et al.*, 2012) e num estudo de produção de fala para a língua germânica (Niebergall *et al.*, 2012).

Maioritariamente, as aquisições em tempo real têm sido efetuadas mediante a realização de um único corte no plano médio-sagital (2D), permitindo obter uma resolução temporal ajustada ao estudo da fala. No entanto, a importância de obter mais informação, para lá da linha média, relevante para algumas classes de sons tem conduzido à implementação de estratégias para obter informação 3D em tempo-real. Vários autores (Bresch *et al.*, 2008b; Proctor *et al.*, 2012; Kim *et al.*, 2012; Zhu *et al.*, 2013a,b) reforçam a importância da visualização 3D dinâmica das características

³⁵Para uma leitor com mais interesse nesta área recomendamos uma visita à página desta equipa: <http://sail.usc.edu/span/publications.php>.

dos articuladores, como sejam, por exemplo, a formação e extensão do *groove* em algumas fricativas, a assimetria na formação dos canais laterais nas consoantes laterais, entre outras.

As primeiras aproximações para obter informação 3D (ao longo do tempo) passaram pela aquisição de mais do que um corte em aquisições separadas. Proctor *et al.* (2012), por exemplo, obtiveram, para além de um corte no plano sagital, um corte no plano coronal, para estudar as consoantes sibilantes do Mandarim. A resolução temporal obtida foi de 18.7 *frames/s* reconstruídas com *sliding windows* para 33 *frames/s*. Outra estratégia muito interessante, todavia complexa e trabalhosa no que diz respeito à reconstrução da informação, é apresentada em Zhu *et al.* (2013a). Estes autores conseguiram uma visualização dinâmica 3D do trato vocal em tempo real, através da aquisição, de forma independente, de cortes parasagittais (21) englobando todo o trato vocal. Cada uma das séries foi obtida com aquisição sincronizada do sinal acústico. Depois de feito o alinhamento de todas as séries, utilizando o sinal acústico como referência, obtiveram reconstruções 3D do trato vocal. A resolução temporal conseguida foi de 78 *ms* e resolução de $2.4 \times 2.4 \times 3 \text{ mm}^3$.

Com sacrifício da resolução temporal, é possível adquirir em simultâneo (na realidade, será virtualmente em simultâneo) vários cortes em planos ortogonais ou paralelos. Um exemplo de aquisição multicorte é apresentado em Kim *et al.* (2011) e Kim *et al.* (2012). Os autores testam dois protocolos: um com dois cortes (sagital e axial), com resolução temporal de 156 *ms*, espessura de corte de 6 *mm*, resolução no plano de $2.4 \times 2.4 \text{ mm}^2$; e outro, com três cortes, obtendo resolução temporal de 163 *ms*, com alguma degradação a nível da resolução espacial ($3 \times 3 \text{ mm}^2$).

A mais recente inovação na área pode já ser considerada uma verdadeira aquisição 3D em tempo real. De facto, Zhu *et al.* (2013b), utilizando uma aquisição com uma range de espirais subamostrada (*3D stack of spirals*), explorando correlações temporo-espaciais (*k-t space*) e um modelo PS (*Partially Separable*) para reconstrução das imagens, conseguiram acelerar substancialmente as aquisições (por um fator de 8), obtendo um volume de $20 \times 20 \times 5 \text{ cm}^3$, com taxa de amostragem de 8.6 *frames/s* e resolução espacial de $3 \times 3 \times 3.1 \text{ mm}^3$. A qualidade das imagens obtidas é, de acordo com os autores, ainda baixa, mas pode ser melhorada e constituirá o caminho a seguir, particularmente no estudo de sons que apresentem maior complexidade geométrica.

Finalmente, uma referência a alguns estudos com RM realizados para várias línguas, uma vez que, apesar de algumas características universais, cada língua apresenta especificidades que importa conhecer e estudar. As línguas Inglesa e Francesa serão aquelas que têm sido mais estudadas, por diferentes equipas de investigação, mas o interesse tem-se alargado a muitas outras línguas (veja-se a tabela 2.4). Caberá ainda salientar que os objetivos subjacentes aos vários estudos que recorrem à RM, na área da produção de fala, são bastante diversificados: aprofundamento de conhecimento na área da fonética e linguística, modelação e síntese, impacto das emoções na fala, estudo da voz cantada, entre muitas outras. Grande parte destes estudos tem sido efetuado em indivíduos voluntários, sem qualquer tipo de alterações. No entanto, a sua aplicação no estudo de variadas patologias, de alguma forma relacionadas ou com impacto na produção de fala, tem merecido o interesse dos investigadores nos últimos anos. Alguns exemplos desses estudos apresentam-se na tabela 2.5.

Tabela 2.4: Exemplos de estudos realizados com RM para várias línguas.

Línguas	Exemplos de Estudos
Inglês	Narayanan <i>et al.</i> (1997a, 2011) Stone e Lundberg (1996), Story <i>et al.</i> (1996), Shadle <i>et al.</i> (1999), Gick <i>et al.</i> (2002)
Francês	Demolin <i>et al.</i> (1996, 2000) Delvaux <i>et al.</i> (2002a), Badin <i>et al.</i> (2002), Engwall <i>et al.</i> (2006)
Alemão	Hoole <i>et al.</i> (2000), Kröger <i>et al.</i> (2000), Birkholz e Kröger (2006)
Sueco	Engwall e Badin (2000), Engwall (2000a), Ericsson (2005)
Russo	Kedrova <i>et al.</i> (2008)
Coreano	Yang (1999), Kim (2004)
Japonês	Takemoto <i>et al.</i> (2006), Takano e Honda (2007)
Italiano	Romano e Badin (2009)
Árabe	Israel <i>et al.</i> (2012), Shosted <i>et al.</i> (2012b)
Mandarim	Proctor <i>et al.</i> (2012)
Tamil	Narayanan <i>et al.</i> (1999)
Português Brasil	Gregio (2006), Medeiros e Demolin (2006)
Português Europeu	Martins (2007), Oliveira (2009), Rua e Freitas (2007)

Tendo em atenção os estudos reportados na área, em particular os mais recentes, poderemos apontar (seguindo a ideia de Masaki *et al.*, 2008) para uma quarta fase de utilização da RM na área da produção de fala. Esta centra-se fundamentalmente na melhoria da resolução temporal e/ou espacial com aquisições de RM em tempo real, permitindo já a obtenção de mais de 30 *frames/s* (em 2D) ou adquirir imagem em tempo real, com verdadeiras aquisições 3D. Tem existido, paralelamente, um grande investimento no desenvolvimento de várias abordagens de processamento dos dados, que se tem revelado fundamental para a obtenção de mais (e diferente) informação a partir das imagens de RM (e.g., relativa à coordenação entre os vários articuladores, aspetos prosódicos). A extrapolação de todo este conhecimento, até aqui maioritariamente obtido em indivíduos sem patologias, para áreas mais clínicas, tem sido crescente.

A fase seguinte passará, tudo indica, por: 1) uma abordagem multimodal com co-registo de informação proveniente de várias técnicas articulatórias como, por exemplo, RM e EMA (Narayanan *et al.*, 2011; Kim *et al.*, 2014), explorando as vantagens e minimizando as limitações de cada uma; 2) pela normalização, quantificação e sistematização na análise dos dados que minimizem a subjetividade na interpretação dos resultados e permitam, por exemplo, lidar com a variabilidade intra e entre falantes (Silva *et al.*, 2013b; Silva e Teixeira, 2014).

A revisão efetuada neste capítulo proporcionou uma visão muito genérica dos estudos que têm sido realizados nesta área, tentando incluir trabalhos conduzidos por diversas equipas de investigação e que recorrem a diferentes técnicas. Relativamente aos estudos estáticos, em Martins (2007) é providenciada uma revisão mais detalhada de muitos dos estudos aqui referenciados. No início dos capítulos 5 e 6, far-se-á um enquadramento mais dirigido para os estudos RM efetuados com consoantes laterais e vogais nasais, uma vez que são estes os sons do PE que se abordam no contexto deste trabalho.

Tabela 2.5: *Estudos de RM na produção e modelação fala - áreas de aplicação.*

Áreas de Aplicação	Exemplos de Estudos
Fonética e linguística	Engwall e Badin (2000), Delvaux <i>et al.</i> (2002a), Martins <i>et al.</i> (2008), Byrd <i>et al.</i> (2009), Lammert <i>et al.</i> (2009, 2010a)
Síntese articulatória	Birkholz <i>et al.</i> (2006), Birkholz (2013) Van Den Doel <i>et al.</i> (2006)
Modelação 3D	Badin <i>et al.</i> (1998, 2002) Engwall (2000a) Serrurier e Badin (2008) Fels <i>et al.</i> (2009)
Estudo do canto	Echternach <i>et al.</i> (2008, 2010) Proctor <i>et al.</i> (2010a) Bresch e Narayanan (2010) Ventura <i>et al.</i> (2013)
Estudo das Emoções	Lee <i>et al.</i> (2006), Lee e Narayanan (2010)
Perturbações articulatórias	Kröger <i>et al.</i> (2005) Hagedorn <i>et al.</i> (2012)
Mordida aberta/mal oclusão	Karacay <i>et al.</i> (2006), Görgülü <i>et al.</i> (2011)
Apneia sono	Okubo <i>et al.</i> (2006), Chan <i>et al.</i> (2010), Kim <i>et al.</i> (2013)
Glossectomizados	Mády <i>et al.</i> (2002), Zhou <i>et al.</i> (2013)
Fendas palatinas e IVF	Drissi <i>et al.</i> (2011), Bae <i>et al.</i> (2011), Perry <i>et al.</i> (2013)
Deglutição	Anagnostara <i>et al.</i> (2001) e Ajaj <i>et al.</i> (2005) Breyer <i>et al.</i> (2009) e Zhang <i>et al.</i> (2012)

Capítulo 3

Processamento de Imagem

One picture is worth a thousand words

OLD FOLK SAYING IN LIANG E LAUTERBUR (2000)

No capítulo precedente, abordámos alguns conceitos e desenvolvimentos técnicos relacionados com a modalidade imagiológica a que se recorre para a realização deste trabalho - a Ressonância Magnética. Para além da aquisição de imagem é necessário efetuar o seu processamento, tendo em vista a obtenção de informação (qualitativa e quantitativa) que permita uma melhor caracterização articulatória dos sons em análise. Desta forma, considerou-se pertinente conferir ao processamento de imagem algum espaço, dentro do que é o enquadramento teórico deste trabalho.

Este trabalho, na área da produção de fala, envolve a obtenção de imagens do trato vocal com recurso a diferentes técnicas de RM. Desta forma, o que está em causa é a análise de estruturas anatómicas e de processos fisiológicos (fala) que, inevitavelmente, nos remetem para uma área específica do processamento de imagem - o processamento de imagem médica. Esta é uma área que tem merecido a atenção e entusiasmo da comunidade científica, nos últimos anos, sendo eminentemente interdisciplinar. Congrega contributos de diferentes áreas do saber (e.g., vários ramos da engenharia) onde as ciências e tecnologias da saúde estão também incluídas (e.g., biomedicina, imagiologia). Apesar de grande parte dos conceitos serem transversais a outras áreas do processamento de imagem, cedo se percebeu que a natureza especial das imagens e das aplicações médicas requer um tratamento específico. A especificidade da imagem médica deve-se aos próprios métodos de aquisição de imagem (e.g., artefactos específicos, restrições de dose) e, também, aos próprios sujeitos em que é adquirida. O facto de lidar com estruturas vivas, que se movem e deformam, que apresentam topologias tridimensionais complexas (variáveis de indivíduo para indivíduo), que se alteram na presença de processos patológicos constitui, a par com a quantidade de dados a processar, um desafio considerável (Bankman, 2000; Sonka e Fitzpatrick, 2004).

O processamento de imagem - termo lato - desenvolve diferentes abordagens a problemas

específicos, com o intuito de realçar a informação contida nos dados brutos das imagens, proporcionando uma visão seletiva e abrindo caminho para posterior análise das imagens. Inclui um espectro bastante largo de tarefas que, para além da identificação das estruturas de interesse (segmentação), engloba processos de melhoria (*enhancement*), co-registo, compressão, armazenamento, entre muitas outras (Sonka e Fitzpatrick, 2004).

No âmbito deste trabalho, foram efetuadas algumas tarefas simples de melhoria para preparação dos dados (e.g., suavização, interpolação) e uma tarefa de co-registo fundamental para alcançar os objetivos propostos (co-registo das arcadas dentárias dos falantes com os volumes obtidos em fonação - ver figura 3.1). A parte substancial do trabalho de processamento foi dedicada a tarefas de segmentação e análise de imagem (obtenção de áreas e volumes, extração de formas, parâmetros), pelo que lhe será conferido um maior destaque.

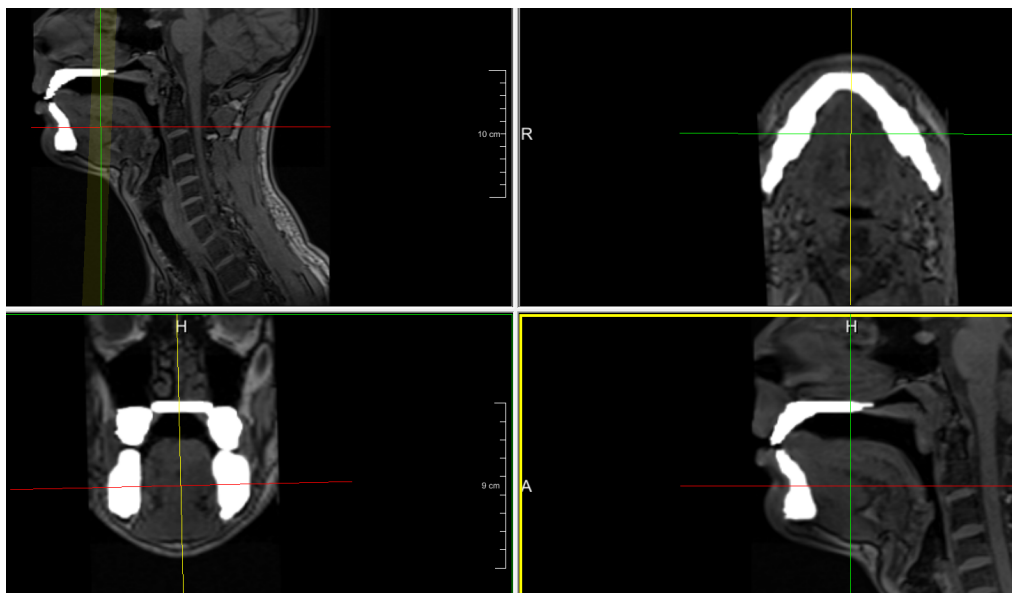


Figura 3.1: Tarefa de co-registo (manual) das arcadas dentárias de um dos falantes participantes no estudo (LCR) com o volume adquirido durante a produção sustentada do /l/ para viabilizar a identificação das passagens laterais.

Assim, neste capítulo, depois de um enquadramento relativo aos métodos de segmentação e às ferramentas de processamento atualmente disponíveis, far-se-á uma síntese das técnicas que têm sido utilizadas em estudos similares, no âmbito da produção de fala. Elencam-se depois os principais problemas associados ao processamento de imagens do trato vocal obtidas durante a produção de fala. No final, apresentam-se as nossas opções quanto aos métodos de segmentação e às ferramentas utilizadas neste trabalho.

3.1 Segmentação de imagem

O principal objetivo da segmentação de imagem é dividir uma imagem em sub-regiões que apresentem propriedades homogêneas (de cor ou textura, por exemplo) às quais correspondem objetos

ou partes destes (Zhu e Yuille, 1996; Pham *et al.*, 2000; Withey e Koles, 2008). De acordo com Falcão *et al.* (2000), o processo de segmentação envolve duas tarefas diferentes, mas intimamente relacionadas: o reconhecimento e a delimitação do objecto. O reconhecimento é o processo que permite, de forma aproximada, identificar a localização do objeto na imagem, permitindo a sua distinção de outros objetos invariavelmente presentes nesta. A delimitação, por sua vez, permite especificar, de forma precisa, a extensão espacial do objeto.

É consensual entre os vários autores consultados que a segmentação de imagem é um processo vital na posterior análise de imagens médicas, permitindo quantificar e visualizar a anatomia e fisiologia subjacentes (Falcão *et al.*, 2000; Hamarneh e Gustavsson, 2000; Hamarneh *et al.*, 2005), orientar o processo de normalização espacial entre diferentes indivíduos e planejar intervenções médicas ou cirúrgicas (Yushkevich *et al.*, 2006).

Nos últimos anos, tem-se observado grandes progressos a nível da segmentação de imagens, particularmente a nível de imagem médica. Apesar disso, a segmentação continua ainda a ser um grande desafio na área das ciências da imagem (Falcão *et al.*, 2000). A segmentação de imagens médicas, na opinião de Withey e Koles (2008), apresenta três problemas principais: 1) as imagens geralmente contêm ruído afetando a intensidade de um determinado *pixel* e dificultando a sua classificação; 2) não existe uniformidade da intensidade ao longo da imagem e 3) as imagens são afetadas por efeitos de volume parcial. Há ainda que considerar que o contraste entre estruturas anatómicas vizinhas pode ser intrinsecamente reduzido ou até virtualmente inexistente, sendo, neste caso, o recurso a informação *a priori* inevitável.

3.1.1 Técnicas de Segmentação de imagem

O conjunto de técnicas de segmentação disponíveis nos dias de hoje é consideravelmente grande. A forma como os métodos são classificados e agrupados é também bastante variável e nem sempre consensual.

3.1.1.1 Nível de interação do utilizador

Uma das formas possíveis e bastante habitual de classificação dos métodos de segmentação tem por base o nível de interação do utilizador. Assim, o leque de técnicas pode variar entre os métodos de segmentação manual e os métodos totalmente automáticos. Entre estes extremos, encontram-se as abordagens através de técnicas semi-automáticas, que conjugam a eficiência e reprodutibilidade da segmentação automática com a capacidade de julgamento e decisão inerentes à experiência humana.

- **Segmentação manual** - Na segmentação manual, os contornos da estrutura a segmentar são efetuados, em geral, por especialistas na área em questão e que são capazes de classificar determinada região como pertencendo, ou não, à estrutura de interesse (Carbone, 2008). A segmentação manual, apesar de permitir um bom reconhecimento do objeto, é um processo extremamente demorado, efetuado corte a corte, ao longo de um volume de dados, podendo

ocupar algumas horas para segmentar uma estrutura (Yushkevich *et al.*, 2006; Poon *et al.*, 2008; Kang, 2005). Exige também algum esforço por parte do operador para obter uma delimitação precisa do objeto (Kang, 2005), podendo conduzir a fadiga e, conseqüentemente, a erros (Yushkevich *et al.*, 2006). Por outro lado, também se tem verificado que existe uma grande variabilidade intra e inter observador (Yushkevich *et al.*, 2006; Poon *et al.*, 2008; Silva, 2005).

- **Segmentação semi-automática** - Devido às dificuldades e limitações apontadas às técnicas manuais, tem existido um considerável interesse no desenvolvimento e implementação de rotinas de segmentação semi-automáticas. Estas conjugam a facilidade na segmentação de estruturas (e.g., em 3D) com uma interação mínima do utilizador (Poon *et al.*, 2008). A intervenção do utilizador é necessária no processo de inicialização e na definição de alguns parâmetros, após o que o contorno evolui ajustando-se à região de interesse. Dentro deste tipo de métodos encontram-se técnicas mais antigas (e.g., *Thresholding*, *Region Growing*) e outras bastante mais recentes e complexas como *Reaction Diffuse Active Contours* (Carbone, 2008).
- **Segmentação automática** - Nos métodos considerados totalmente automáticos, no extremo oposto do espectro das técnicas de segmentação, o algoritmo efetua a segmentação, tendo por base um conjunto de regras pré-definidas. São baseadas em modelos probabilísticos de intensidade de imagem, deformação ou modelos estatísticos de forma (Yushkevich *et al.*, 2006). Apesar destas técnicas serem bastante promissoras, constituem, ainda, um problema não totalmente resolvido. No caso particular da imagem médica, terá que existir uma sintonia precisa entre os parâmetros de segmentação e as estruturas anatómicas, que é complicada com a diversidade entre sujeitos, com a variação à norma e existência de patologias associadas (Poon *et al.*, 2008). Funcionam muito bem, sendo métodos robustos, consistentes e eficientes, quando há um ajuste perfeito dos parâmetros, o que nem sempre se consegue.

De acordo com Yushkevich *et al.* (2006) e apesar do vasto leque de técnicas atualmente existentes, ainda são as técnicas de segmentação manual as mais utilizadas para efetuar, por exemplo, a segmentação de imagens médicas. Na opinião destes autores, no caso das técnicas totalmente automáticas, a razão para a sua pouca utilização poderá estar relacionada com a sua insuficiente exatidão na delimitação de estruturas anatómicas, quando estas se afastam da norma, quando existe patologia associada e, ainda, devido às exigências computacionais requeridas por estes métodos. Por outro lado, a fraca adesão aos métodos semi-automáticos ficar-se-á a dever à falta de interfaces simples e intuitivas para um utilizador que não tenha conhecimentos de programação (Yushkevich *et al.*, 2006). Note-se, no entanto, que, nos últimos anos, têm surgido plataformas bastante interessantes, que implementam vários métodos de segmentação, para utilizadores sem este tipo de conhecimento.

3.1.1.2 Outras classificações

A divisão dos métodos de segmentação que apresentámos teve como base o nível de interação do operador. Na bibliografia consultada, encontram-se outras formas de classificação destas técnicas. Por exemplo, Pham *et al.* (2000), numa revisão acerca dos métodos de segmentação, consideram oito categorias: os métodos baseados em *threshold* ou limiarização, *Region Growing*, abordagens com classificadores, métodos de *clustering*, modelos de *Markov* (*Markov Random Field*), redes neurais (*Artificial neural networks*), modelos deformáveis e métodos baseados em atlas. Outros, como Withey e Koles (2008), efetuam uma divisão dos métodos de segmentação, considerando várias gerações, cada uma das quais representando diferentes níveis de desenvolvimento algorítmico.

Uma descrição abrangente e detalhada dos vários métodos de segmentação e da sua classificação pode ser encontrada em vários capítulos das obras editadas por Bankman (2000), Sonka e Fitzpatrick (2004), Sonka *et al.* (2007) e também nos artigos de revisão já referidos como Withey e Koles (2008) ou Pham *et al.* (2000). Nesta revisão, conferimos um pouco mais de destaque aos métodos a que recorreremos no decurso do nosso trabalho para efetuar o processamento das imagens, nomeadamente alguns dos métodos incluídos dentro da classe dos modelos deformáveis e o *Region Growing*.

3.1.2 Modelos deformáveis

Os modelos deformáveis são bastante apelativos para processamento de imagem médica, uma vez que lidam bastante bem com a complexidade geométrica, deformação e variabilidade anatômica que está associada às estruturas biológicas (McInerney e Terzopoulos, 1996). Para reconhecer e delinear estas estruturas ou acompanhar o seu movimento, os métodos de segmentação baseados nestes modelos contam com informação contida nas próprias imagens e com incorporação de conhecimento prévio relativo à localização, dimensão ou forma das estruturas. O conhecimento do próprio operador pode também ser integrado, no sentido de melhor guiar o processo de segmentação.

Estes modelos incluem uma grande variedade de métodos de segmentação. No que diz respeito à terminologia, vários nomes têm sido utilizados para referir os modelos deformáveis como contornos ou superfícies ativas, *snakes*, *balloons*, entre outros.

Na figura 3.2, apresenta-se um esquema que, de forma simplificada e bastante genérica, pode albergar os métodos de segmentação baseados em modelos deformáveis. Todos os modelos deformáveis partem de um modelo inicial, flexível, que mediante um processo de evolução (ou delineação, ou correspondência) - descrito por um modelo matemático dinâmico - é iterativamente orientado em direção a características relevantes da imagem, permitindo obter um contorno final.

Basicamente, o que diverge nos vários métodos são os modelos matemáticos que lhes estão subjacentes e o tipo de informação *a priori* que tem que ser definida, à entrada do processo, para conduzir a evolução. Assim, este processo pode ser governado por equações diferenciais parciais (e.g., *snakes*, *level-sets*), por modelos estatísticos (*Active Appearance Models* - AAM ou *Active Shape Models* - ASM) ou por grafos (e.g., *live-wire*).

Nas *snakes* e nos *level-sets*, ambos são formulados através de equações diferenciais, mas recorrem a termos diferentes para as resolver. As *snakes* são parametrizadas e, por isso, designadas por modelos deformáveis paramétricos, enquanto que os *level sets* se servem de características geométricas do objeto para resolver as equações.

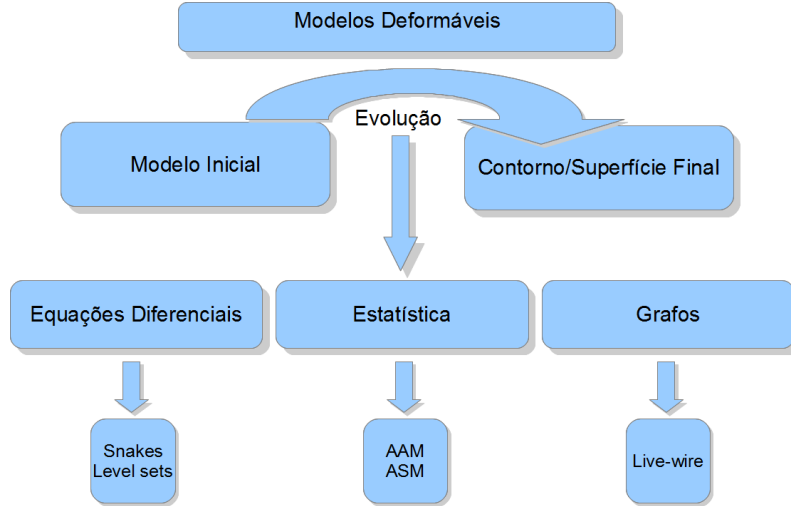


Figura 3.2: Taxonomia relativa aos modelos deformáveis.

3.1.2.1 *Snakes* - Modelos paramétricos

As *snakes*, também chamadas curvas de minimização de energia, pertencem, como já referido, à classe dos modelos deformáveis paramétricos e foram introduzidos por Kass *et al.*, em 1987. Desde aí, têm sido amplamente utilizadas para análise de imagens. Uma área em que a utilização das *snakes* se tem mostrado particularmente relevante e bastante popular é, precisamente, no âmbito da imagem médica/biomédica e, também, no seguimento (*tracking*) de objetos em imagens de vídeo (Liang *et al.*, 2006).

Na sua formulação clássica, as *snakes* podem ser consideradas como contornos ou curvas definidas no domínio da imagem, que se deformam, visando atingir um estado de energia mínima (*local minimum*). A curva $C(s)$ é parametrizada por $s \in [0, 1]$, sendo a energia da *snake* definida, matematicamente, como:

$$E(C) = \int_0^1 (E_{int} + E_{im} + E_{user}) ds \quad (3.1)$$

onde E_{int} representa a energia interna e que, tipicamente, inclui termos relativos à elasticidade e à rigidez; E_{im} representa a energia da imagem muitas vezes chamada função de energia potencial - atrai o contorno para as arestas ou contornos desejados em função de magnitude do gradiente; E_{user} é um termo que contempla a imposição de constrangimentos pelo operador (inicialização) ou introdução de conhecimento prévio. A *snake* deforma-se minimizando esta energia para obter a delineação do objeto (Niessen *et al.*, 1998).

Apesar de muito utilizadas, são conhecidas algumas limitações (Niessen *et al.*, 1998; Xu

et al., 2000; Silva *et al.*, 2004) das *snakes*, nomeadamente:

1. os problemas com a inicialização - as *snakes* funcionam relativamente bem, se a inicialização é efetuada próxima do contorno desejado;
2. a sua falta de flexibilidade devido a uma parametrização fixa - nas situações que o modelo inicial e o contorno diferem muito no tamanho e forma, o modelo terá que ser re-parametrizado de forma dinâmica. Em 2D, o processo é relativamente simples mas em 3D requer soluções complexas do ponto de vista computacional;
3. a sua incapacidade em lidar com a topologia do objeto - esta tem que ser conhecida *a priori*.

Contudo, desde a sua formulação inicial, várias abordagens e variantes (e.g., *T-snakes*) foram desenvolvidas com o intuito de orientar o processo de evolução, considerando critérios como a tolerância na inicialização, a velocidade de convergência, a variação topológica, a sensibilidade ao ruído e a introdução de conhecimento externo. Por exemplo, as sucessivas alterações propostas por Cohen *et al.* (1992) e Cohen e Cohen (1993) conduziram a melhorias, tornando as *snakes* menos sensíveis ao processo de inicialização. Estes autores incluem uma força de inflação para tornar o método menos sensível a “falsas” arestas.

Region Competition Snakes

Ainda dentro do domínio das *snakes*, Zhu e Yuille (1996) propuseram um método de segmentação mais robusto designado por *Region Competition Snakes*. Constitui uma abordagem elaborada em que os méritos das *snakes* tradicionais e as características estatísticas do *Region growing* são congregados através de uma descrição variacional onde o processo de otimização é conseguido através de um critério *Bayesiano* (*Minimum Description Length* - *MDL*). Este é um dos métodos de segmentação implementado no *ITK-SNAP*, que referiremos mais adiante.

Tal como nas *snakes*, em geral, as forças internas derivam da geometria do contorno e são utilizadas para impor constrangimentos. A força externa é calculada, estimando a probabilidade de um determinado *voxel*, do conjunto de *voxels* da imagem, pertencer a cada uma das regiões em competição. Na proposta inicial de Zhu e Yuille (1996), é calculada a força externa, tendo como base as diferenças logarítmicas das duas probabilidades, enquanto que no *ITK-SNAP* as probabilidades são estimadas utilizando *Fuzzy Thresholds* (Yushkevich *et al.*, 2006).

3.1.2.2 *Level-sets* - Modelos geométricos

Os modelos geométricos deformáveis também denominados por *Implicit snakes* foram apresentados, de forma independente, por Caselles *et al.* (1997) e por Malladi *et al.* (1995) sob a designação, respetivamente, de *Geodesic Active Contours* (Contornos ativos Geodésicos) e *Level-set Front Propagation Models*, tendo como base o trabalho de Osher e Sethian (1988).

Esta abordagem posiciona-se como uma alternativa à descrição dos contornos ativos paramétricos. A estimativa da evolução da curva é efetuada através da resolução numérica das equações

diferenciais parciais, mas orientada por constrangimentos geométricos tais como: curvatura, gradiente ou outros. As curvas ou superfícies alvo são integradas em funções de nível mais elevado e, conseqüentemente, a evolução é orientada, fazendo corresponder um determinado *level-set* ou curva de nível a características relevantes na imagem (Niessen *et al.*, 1998). Matematicamente, de acordo com a mesma fonte, a equação parcial pode ser representada por:

$$\frac{\partial I}{\partial t} = g(\alpha I_w + I_{vv}) \quad (3.2)$$

onde v e w representam, respetivamente, a direção tangencial e normal à curva, α é uma constante e $g(x, y, \sigma)$ é a função de velocidade que é tipicamente uma função decrescente do gradiente de imagem ∇L à escala σ . Se I satisfizer a condição de $\nabla I \neq 0$ para o nível zero da função C , a evolução de C será dada por:

$$\frac{\partial C}{\partial t} = g(\alpha + k) \vec{N} \quad (3.3)$$

O suporte dos *level-set* permite uma segmentação perfeita de vários objetos, superando também o problema da variação topológica ao longo de todo o conjunto de imagens, que constitui uma limitação das *snakes* parametrizadas (Niessen *et al.*, 1998).

No entanto, apesar de matematicamente constituírem uma plataforma atrativa, no que diz respeito às equações parciais diferenciais, estas podem não ser tão convenientes como as variantes parametrizadas, quando um utilizador pretende incorporar, de forma interativa, mecanismos de controlo adicional relacionados com a evolução da deformação. Para além disso, a existência de uma função de dimensão elevada dificulta a imposição de constrangimentos geométricos ou topológicos nas respetivas curvas de nível. Assim, esta formulação implícita da evolução dos contornos ou superfícies pode dificultar a sua utilização, afetando a sua eficiência e o grau de automatização no processo de segmentação (McInerney e Terzopoulos, 1996).

Uma visão genérica global acerca dos modelos geométricos deformáveis pode ser encontrada em Sonka *et al.* (2007) e um tratamento mais aprofundado é proporcionado por Osher e Paragios (2003) e Osher e Sethian (1988).

3.1.2.3 *Live-wire* ou *Intelligent Scissors*

As técnicas *Live*, que incluem o *live-wire* (2D e 3D) e *live-lane* foram apresentadas de forma independente por Falcão (1997) e por Mortensen e Barrett (1995). Estes últimos autores designam esta técnica como *Intelligent Scissors*. O *live-wire* é um método de segmentação semi-automático, orientado a contornos, conduzido por informação contida na imagem. Esta informação pode ser o gradiente da imagem, zeros laplacianos, valores de intensidade ou outra (Schenk *et al.*, 2000; Salah *et al.*, 2005). Segundo Falcão (1997), é um método que permite um grande controlo do processo de segmentação, por parte do utilizador (com diferentes níveis de interatividade), tendo a capacidade de resolver problemas de segmentação, quando métodos mais automatizados não se adequam ou falham no processo de segmentação. Reduz o esforço e tempo de segmentação na obtenção de um

contorno quando comparado com a delineação manual do mesmo.

Neste método, a obtenção dos contornos de um objeto é baseado no cálculo dos percursos de custo mínimo entre os vértices de um grafo¹. Na figura 3.3, apresenta-se um diagrama que representa um grafo. Habitualmente, representa-se um grafo simples por um diagrama no qual os vértices são representados por pontos (a, b, c, \dots) e as arestas por linhas, unindo vértices adjacentes (Picado, 2010 p.54). Numa imagem, os vértices correspondem a *pixels* e as arestas estabelecem a conexão entre os *pixels* vizinhos (Schenk *et al.*, 2000).

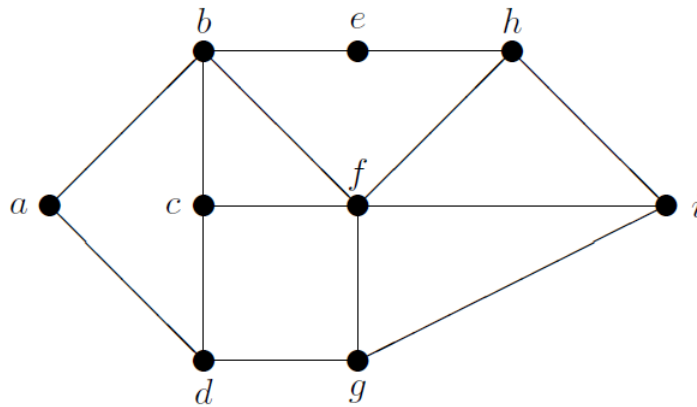


Figura 3.3: Diagrama que representa um grafo. Fonte da imagem: Picado (2010).

O peso das arestas é definido em termos de funções de custo local. Os diferentes caminhos podem ser calculados através do algoritmo de procura de *Dijkstra* e a função de custo local é representada pela soma ponderada dos diferentes componentes funcionais que descrevem as características da aresta. A extração do contorno é efetuada iterativamente, colocando um ponto semente na fronteira do objeto de interesse. Quando se coloca um segundo ponto é criada uma via de custo mínimo que, idealmente, vai aderir à fronteira do objeto (Schenk *et al.*, 2000; Salah *et al.*, 2005).

O *live-wire* é um método intrinsecamente 2D, no entanto pode ser estendido ao processamento de volumes, através da segmentação sucessiva de cortes 2D ao longo de um volume. Na figura 3.4, apresenta-se um exemplo de segmentação do trato vocal, utilizando *live-wire*, ao longo de uma série de cortes pertencentes a um volume de dados, permitindo reconstruir uma superfície que representa o objeto 3D. Este processo foi considerado por Carbone (2008) como uma segmentação 2,5D. No entanto, alguns autores consideram esta abordagem, para segmentação de volumes, pouco eficiente e entediante (Hamarnah *et al.*, 2005; Salah *et al.*, 2005; Poon *et al.*, 2007).

Uma extensão do conceito *live-wire* de 2D para 3D, proporcionando uma maior eficiência, foi proposto por Hamarnah *et al.* (2005) e, posteriormente, modificado e evoluído por Poon *et al.* (2007, 2008) para albergar a segmentação de objetos com topologia mais complexa (e.g., geometria

¹Um grafo simples G é um conjunto finito e não vazio $V(G)$ de elementos chamados vértices e um conjunto finito $A(G)$ de pares não ordenados de elementos distintos de $V(G)$, chamados arestas. Definição apresentada em Picado (2010), p.53.

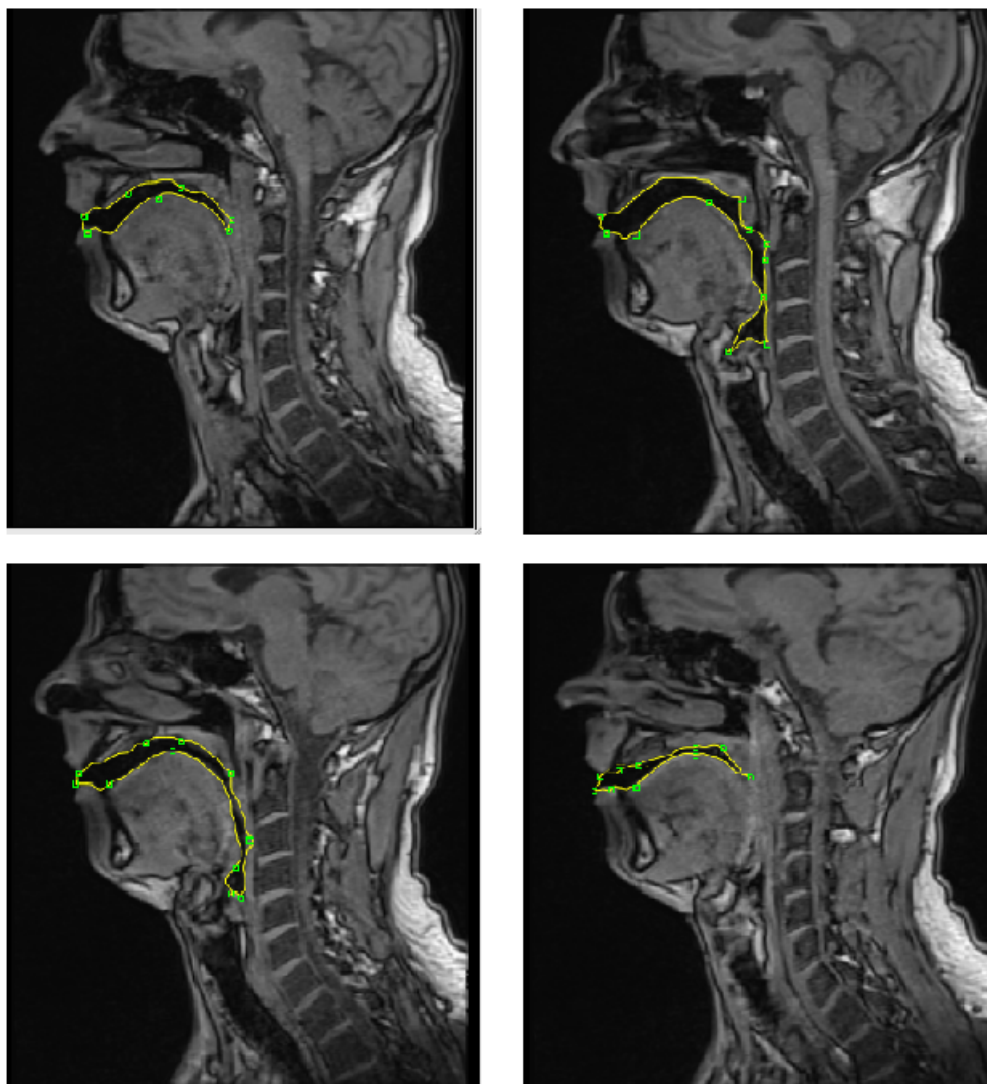


Figura 3.4: Segmentação do trato vocal em imagens sagitais, utilizando *live-wire* ao longo de uma série de cortes 2D pertencentes a um volume de dados adquirido no âmbito deste trabalho.

não esférica, concavidades, protruções).

O processo inicia-se efetuando segmentação 2D *live-wire*, em cortes separados, em dois planos ortogonais. Estes contornos vão servir para a imposição de constrangimentos na segmentação global. Constituem os pontos semente que serão utilizados na terceira orientação (ou, plano), intercetando os contornos 2D efetuados e os cortes não segmentados. Os pontos de interceção são pré-processados e usados para criar um mapa em “tartaruga” (*turtle map*), que representa um conjunto de segmentos de linha ortogonais. É aplicado depois um algoritmo de ordenação, por forma a que os pontos ordenados imitem a sequência de pontos que o utilizador deve selecionar durante a segmentação, mas agora de forma completamente automática nos cortes não segmentados (Poon *et al.*, 2007, 2008). De acordo com estes autores, o método demonstrou ser bastante mais eficiente do que a segmentação com *live-wire 2D* (realização da tarefa em cerca de 20% do tempo), reproduzível (variou entre 93.2 a 98.7%, nas várias estruturas segmentadas) e exato (quando comparado com traçados manuais) na segmentação de objetos sintéticos, mas também em estruturas anatómicas como o lobo caudado e os ventrículos laterais. A figura 3.5 mostra a implementação do *live-wire 3D*, proposta por Poon *et al.* (2008).

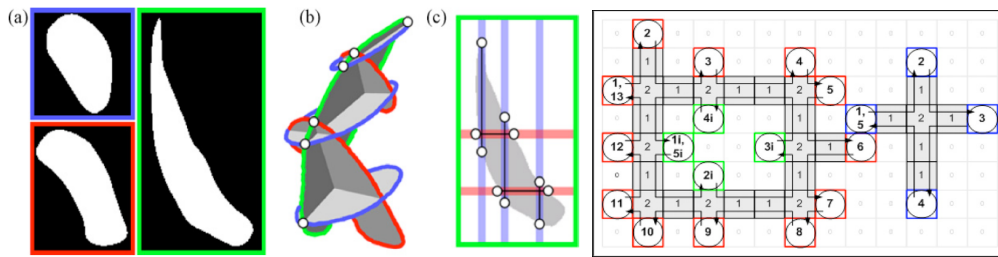


Figura 3.5: *Live-wire 3D* para segmentação de objetos com topologias complexas. Fonte da imagem: Poon *et al.* (2008).

Outras abordagens para segmentação 3D com o *live-wire* podem ser encontradas na literatura como as propostas apresentadas por Salah *et al.* (2005) - método de propagação adaptativa - e por Schenk *et al.* (2000), onde se propõe a utilização do *live-wire* em conjunto com algoritmos de interpolação baseados na forma (*Shape-based interpolation*).

No primeiro caso, os contornos nos cortes não segmentados são gerados (não interpolados), sem requerer intervenção significativa do operador. A ideia nuclear é propagar pontos de controlo através dos cortes e, nesses, a posição dos pontos ser automaticamente ajustada.

No método proposto por Schenk *et al.* (2000), que é a aplicação implementada no *Me-VisLab* para a segmentação de volumes, o operador efetua segmentação com o *live-wire* em determinados cortes, depois os contornos nos cortes intermédios são calculados, de forma automática, recorrendo a um algoritmo de interpolação baseado na forma. O primeiro e último cortes do objeto de interesse têm que ser obrigatoriamente segmentados com o *live-wire*. O processo de interpolação é depois otimizado: partindo do primeiro ponto do contorno, o algoritmo procura *pixels* vizinhos, na direção contrária à dos ponteiros do relógio, até o primeiro ponto ser de novo alcançado. O processo é repetido até que todos os pontos da fronteira sejam atribuídos a uma via.

Os métodos baseados em *snakes* e o *live-wire* partilham algumas semelhanças (e.g., os contornos são atraídos em direção a arestas fortes e requerem a interação do utilizador no processo de segmentação), mas existem algumas diferenças importantes como, por exemplo, o facto de as *snakes* produzirem uma única solução para o problema que pode não se ajustar ao contorno pretendido, enquanto que o *live-wire* apresenta várias alternativas de custo mínimo que o operador pode seleccionar, existindo, desta forma, um maior controlo e previsão sobre o processo de segmentação. Remetemos para Mortensen e Barrett (1995), onde se pode encontrar uma comparação aprofundada entre estas duas técnicas.

Alguns autores, como Liang *et al.* (2006), consideram as técnicas *live-wire* como um caso particular das *snakes*. Ao invés de as considerarem técnicas em competição, os autores defendem que as duas técnicas podem ser conjugadas, tirando partido do melhor que cada uma proporciona. Desta forma, propõem uma plataforma relativamente simples, designada como *United Snakes* que combina vários tipos de *snakes* com *live-wire*, proporcionando uma abordagem de segmentação interativa, com menos intervenção do utilizador e com controlo mais flexível. Esta aplicação tem sido utilizada em várias tarefas no domínio do processamento de imagem médica como, por exemplo, na segmentação de dendrites neuronais ou na quantificação do crescimento de placas em pacientes com esclerose múltipla, mostrando-se uma técnica robusta (veja-se Liang *et al.*, 2006).

3.1.3 *Region Growing*

O método conhecido como *Region Growing* (RG) é uma rotina de segmentação semi-automática, orientada a regiões, tal como o próprio nome indica. A filosofia subjacente a todas as variantes do RG é que o processo se inicia a partir de um ou de vários elementos da imagem (*pixel* ou *voxel*), dando origem a uma região que vai crescendo, iterativamente, até incorporar todos os elementos, na sua vizinhança, que apresentem características (e.g., intensidade do *pixel*) ou critérios de semelhança com o elemento inicial (Lin *et al.*, 2000; Wirjadi, 2007; del Fresno *et al.*, 2009).

É considerado um método de segmentação simples, flexível e robusto, permitindo a delimitação de regiões de interesse a partir da semente inicial (*seed*) colocada dentro de cada uma das regiões a segmentar. O *output* é um conjunto de *pixels* ou *voxels* conectados e que representam a região de interesse (del Fresno *et al.*, 2009). Este método, com colocação de sementes (*Seeded region growing*), foi proposto por Adams e Bischof (1994) e requer a intervenção de um utilizador para a colocação de uma ou de várias sementes (para inicialização do processo), sem que haja necessidade de escolha de outros parâmetros. O número de sementes a colocar é variável, verificando-se que, usualmente, são necessárias mais sementes, se a geometria do objeto é complexa, ou se a gama de contrastes na imagem é grande.

O método é frequentemente sensível ao ruído presente nas imagens, não lida adequadamente com mudanças de tipologia (e.g., descontinuidades), pode haver dificuldade no processo de escolha inicial da semente e na definição do critério de paragem (Carbone, 2008; Lin *et al.*, 2000). Para eliminar a dependência da intervenção de um utilizador, existem já algumas propostas que permitem a colocação de sementes de forma automática (Lin *et al.*, 2000). Tal como nos métodos de limiarização, muitas vezes, o RG não é utilizado sozinho, mas em conjunto com outros métodos

de segmentação (Pham *et al.*, 2000; del Fresno *et al.*, 2009).

3.1.4 Ferramentas atualmente disponíveis

Como já referimos, a área do processamento de imagem médica tem merecido grande atenção e entusiasmo nos últimos anos. Gradualmente, têm vindo a ser desenvolvidos vários pacotes de *software* dedicado, para segmentação e análise de imagem médica, que são colocados à disposição da comunidade científica com interesse neste domínio. Muitos são de acesso livre, com permissão de uso de todas as funcionalidades ou, pelo menos, com acesso a alguns módulos ou versões. Algumas destas ferramentas foram desenvolvidas com o intuito de dar suporte à área da neuro-imagem, outras são mais abrangentes.

Numa compilação efetuada por Withey e Koles (2008), são referidas algumas dessas ferramentas de acesso livre, das quais selecionamos alguns exemplos:

- *BIC Software toolbox*, disponibilizado pelo *McConnell Brain Imaging Centre of the Montreal Neurological Institute*;
- *BrainSuite*, resulta de um projeto conjunto do *Laboratory of Neuro Imaging, University of California Los Angeles (UCLA)* e do *Biomedical Imaging Research Lab of the Signal and Imaging Processing Institute - USC*;
- *Insight Segmentation and Registration Toolkit (ITK)* e o *Visualization Toolkit (VTK)*, desenvolvidos pela *USA National Library of Medicine*, como suporte ao *Visible Human Project*;
- *ITK-SNAP*, que descreveremos, mais adiante, neste capítulo;
- *Analyse*, apresentado pelo *Biomedical Imaging Resource of the Mayo Foundation - Rochester*;
- *3D Slicer*, desenvolvido pelo *MIT Artificial Intelligence Lab and the Surgical Planning Lab at Brigham and Women's Hospital*.

Para além destas, acrescentamos duas outras ferramentas: uma utilizada neste trabalho (*MeVisLab*) e outra que nos foi proporcionada uma apresentação e demonstração - o *Mimics* (desenvolvido pela *Materialise* - <http://biomedical.materialise.com/mimics>) sendo especialmente vocacionada para segmentação e análise de imagens 3D. Esta última ferramenta não é de acesso livre, mas permite a utilização de uma versão de demonstração por período limitado.

Uma análise bastante completa relativa a várias ferramentas de processamento e suas potencialidades é efetuada por Silva (2012). Uma lista detalhada de *software* é disponibilizada através do *Internet Analysis Tools Registry* (<http://cma.mgh.harvard.edu/iatr/>).

3.2 Segmentação de imagem nos estudos de produção de fala

Uma vez que este trabalho se enquadra no âmbito dos estudos de produção de fala, será importante, neste momento, fazer uma síntese relativa ao que tem sido a utilização dos métodos de segmentação de imagem na área. Como seria impossível enumerar todos os estudos, contentar-nos-emos em referir alguns, mais relacionados com o nosso trabalho. Por uma questão de sistematização, incluiremos num grupo os estudos estáticos (produções sustidas), que envolvem a aquisição e processamento de imagens 2D e 3D e, noutro grupo, as aquisições dinâmicas, em particular, os estudos com aquisição de imagem em tempo real.

3.2.1 Estudos envolvendo aquisições 2D e 3D estáticas

Uma das grandes vantagens da RM no contexto da produção de fala é permitir observar o trato vocal em toda a sua extensão. Não menos importante é a possibilidade que oferece para modelação tridimensional das várias estruturas envolvidas. Assim, grande parte dos estudos na área inclui, para além de aquisições 2D (permitindo obter contornos e distâncias médio sagitais), a obtenção de informação 3D a partir de uma série de cortes 2D ou, menos frequentemente, a partir de verdadeiras aquisições volumétricas.

Dependendo dos objetivos de cada estudo, é habitual a necessidade de, por exemplo, extrair as funções de área do trato vocal (interesse na modelação acústico-articulatória e síntese articulatória); obter *renderings* ou visualizações 3D² das diferentes estruturas (e.g., observar características como *grooving* e compressão lateral língua, forma e postura assumida por um articulador); modelação 3D de vários articuladores como a língua, o véu palatino, os lábios (e.g., para construção de modelos biomecânicos ou geométricos). Para obter esta informação, a partir das imagens, têm sido utilizadas várias abordagens.

O recurso a técnicas de segmentação manual é referido em vários estudos mais antigos (Greenwood *et al.*, 1992; Engwall, 2000a; Badin *et al.*, 2000; Kröger *et al.*, 2000; Soquet *et al.*, 2002; Badin *et al.*, 2002), mas também em estudos relativamente mais recentes como:

- Kröger *et al.* (2005) - para obter contornos de todos os articuladores (lábios, véu palatino, palato, língua) a partir de imagens 2D no plano sagital para um falante da língua germânica (vogais longas, lateral, nasais, fricativas);
- Birkholz *et al.* (2006) - com intuito de construir um modelo 3D do trato vocal a partir de imagens de RM. Inclui volumes (18 imagens sagitais) e imagem dinâmicas. Neste estudo, é utilizada segmentação manual com auxílio de um detetor de arestas;

²De acordo com uma revisão do tema efetuada por Carbone (2008) existe, na literatura consultada, uma utilização indiferenciada dos termos visualização 3D e modelação 3D. Assim, por vezes chama-se modelo 3D ao que não é mais que uma visualização 3D. Clarifique-se que na criação de modelos são extraídos parâmetros que podem depois ser manipulados para simular diferentes situações, enquanto que numa visualização apenas se apresentam dados ou reconstruções destes - veja-se Carbone (2008), pág.61.

- Gérard *et al.* (2006) - visando a construção de um modelo biomecânico da língua para estudar produção de fala;
- Takano e Honda (2007) - para segmentação dos músculos da língua, no sentido de avaliar a função dos músculos extrínsecos da língua na produção de vogais, em falantes da língua japonesa;
- Serrurier e Badin (2005a, 2008) - para a construção de um modelo articulatorio 3D do véu palatino a partir de imagens de TC e RM;
- Ventura *et al.* (2008) - com o intuito de obter modelos 3D da língua durante a produção de fala. Segmentação manual efetuada em imagens coronais, sagitais e axiais, com interpolação (curvas de *Bézier*).

O recurso a técnicas de *threshold* para efetuar a segmentação das imagens é várias vezes referido. Baer *et al.* (1991) utilizam técnica de *thresholding* seguida de verificação manual. A mesma técnica é referenciada noutros estudos articulatorios realizados para o inglês americano, em que se visava efetuar caracterização articulatoria detalhada de sons como as vogais, as laterais e os róticos em bases de dados consideravelmente extensas (e.g., Narayanan *et al.*, 1997a). A utilização de *threshold* com limiar fixo é referido, por exemplo, por Badin *et al.* (1998), Engwall e Badin (1999) e Engwall (2003).

O *Region growing* 2D ou 3D tem sido uma das técnicas mais utilizadas no contexto dos estudos de produção de fala, quer em estudos mais antigos (e.g., Story *et al.*, 1996; Behrends e Wismüller, 2001 ou Takemoto *et al.*, 2003), quer em estudos bastante mais recentes:

- Martins *et al.* (2008) e Carbone (2008) - utilizam *Region growing* 2D para segmentação de imagens estáticas numa base de dados recolhida para o PE. De forma exploratória, foi testado para segmentação de imagens em tempo real (obtidas durante a produção de fala);
- Wismueller *et al.* (2008) - recorrem a *Region growing* 3D para segmentação do trato vocal em imagens estáticas. Para prevenir a evolução da segmentação, para o exterior do trato vocal, quando a boca está aberta, os autores referem a necessidade de utilizar um *Kernel* adequado (*I-shaped*);
- Bone *et al.* (2011) - referem a utilização de *Region growing* 3D para segmentação de imagens relativas a volumes da língua, num estudo articulatorio com vogais altas, sibilantes e consoantes líquidas (ver figura 3.6). As malhas de superfície obtidas permitiram, segundo os autores, o estudo da articulação e avaliação da variabilidade entre os falantes.

Nos estudos de produção de fala, o recurso a técnicas baseadas em modelos deformáveis (e.g., como *snakes* e *live-wire*) não tem sido tão frequente. A maior parte dos estudos que recorre a contornos ativos, com exceção do estudo de Demolin *et al.* (1998) que já refere a utilização de *snakes*, é bastante recente³.

³Alguns factos poderão estar na origem deste interesse, mais recente, nos contornos ativos: 1) apesar de serem

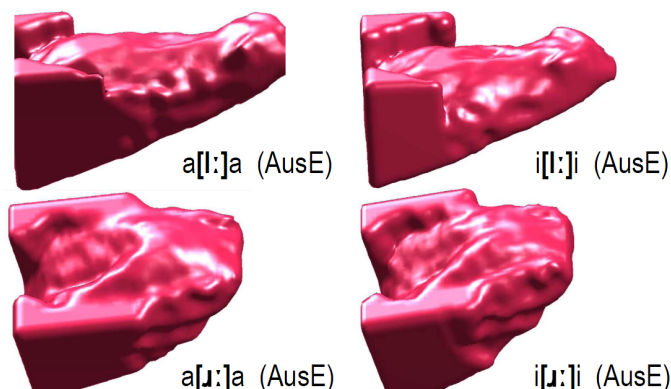


Figura 3.6: Visualizações 3D da língua obtidas com *Region Growing*. Fonte da imagem: Bone *et al.* (2011).

Martins *et al.*, em 2008 recorrem a *live-wire* para segmentação do trato vocal, a partir de aquisições 3D, para vários sons do PE. A abordagem foi implementada por Carbone (2008), que a considerou uma segmentação 2.5D. Permitiu obter áreas, funções de área e visualizações 3D do trato vocal numa base dados bastante extensa.

O recurso ao *live-wire* 2D e 3D tem vindo a ser reportado para segmentação de imagens de alta resolução (e.g., língua, véu palatino), no contexto de um projeto (Dynamic Modeling of the Oral, Pharyngeal, and Laryngeal Complex for Biomedical Applications - OPAL), que visa a construção de modelos biomecânicos de estruturas orais e faríngeas para futura integração com um sintetizador articulatório (ArtiSynth), desenvolvido pela *University of British Columbia*. Um exemplo de utilização de *live-wire* 3D para segmentação do véu palatino (veja-se a figura 3.7) é apresentado em Fels *et al.* (2009), no contexto desse projeto. Para a segmentação das vias aéreas, os autores referem a utilização de outros métodos baseados em contornos ativos como os *Chan Veese active contours*.

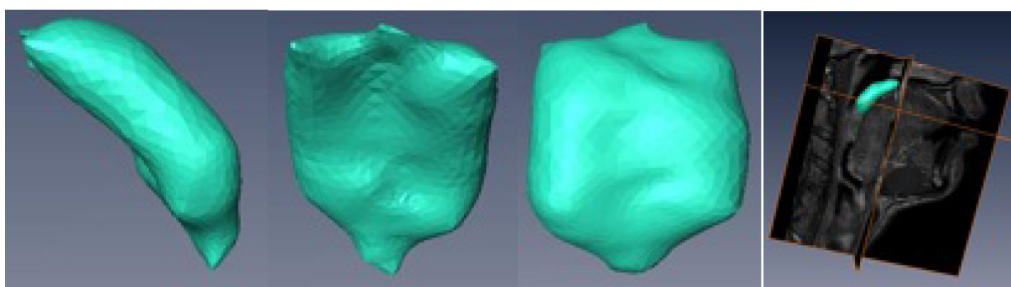


Figura 3.7: Utilização de *Live-wire 3D* para segmentação do véu palatino. Fonte da imagem: Fels *et al.* (2009).

Vogt *et al.* (2009) apresentam segmentação de imagens da língua, recorrendo a uma implementação com a plataforma *ITK*, cuja *pipeline* de segmentação se desenvolve em dois passos:

técnicas propostas há vários anos (anos 90), a sua aplicação na área médica ou biomédica é mais recente; 2) a evolução nas técnicas de aquisição (com RM) tem proporcionado uma melhoria na qualidade dos volumes obtidos o que veio abrir a possibilidade de utilizar outras abordagens de segmentação (e.g., 3D balloons ou bubbles); 3) a maior disponibilização de plataformas de processamento de imagem, intuitivas para o utilizador, que integram métodos baseados em contornos ativos.

primeiro, é utilizado um contorno geodésico ativo para obter uma segmentação preliminar; posteriormente, essa segmentação é refinada com recurso a *level-set* laplaciano. De acordo com os autores, o processo é ainda bastante demorado, falha em algumas regiões da língua, necessitando ser aperfeiçoado. Na imagem da figura 3.8, pode observar-se o resultado da segmentação obtida referente à produção de um som rótico.

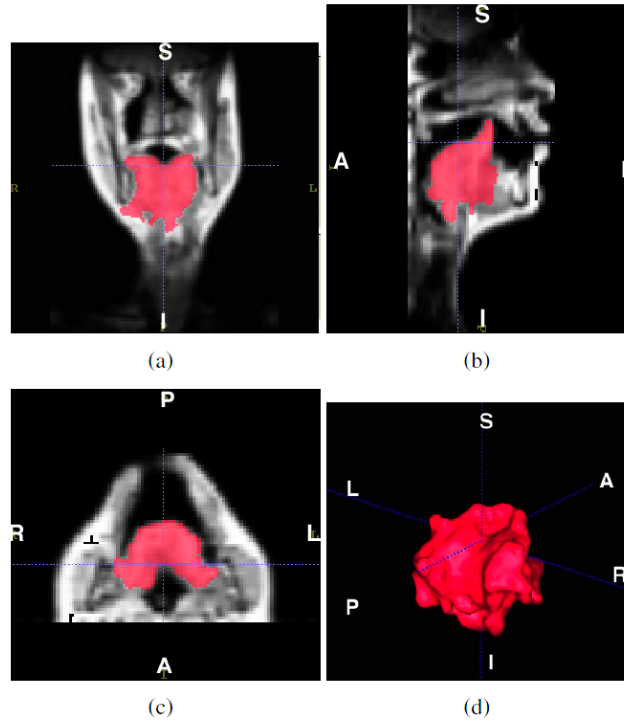


Figura 3.8: Segmentação com contornos geodésicos ativos refinada com recurso a *level-set* laplaciano. Imagem nos três planos referente à segmentação da língua para um som rótico. Fonte de imagem: Vogt *et al.* (2009).

Suo *et al.* (2011) exploram a utilização de um método semi-automático (*Random Walker*) para a segmentação simultânea de várias estruturas do trato vocal (articuladores), em imagens 2D, 2D *cine* e 3D. Este método é, à semelhança do *live-wire*, um algoritmo de procura em grafos. É necessária a intervenção de um utilizador que começa por selecionar *pixels* semente que pertencem a diferentes classes (K). A partir daí, para cada *pixel* não selecionado, o algoritmo procura a probabilidade de um *random walker*, que tenha início nesse local, atingir cada um dos *pixels* semente. Assim, cada um dos *pixels* não selecionados, é atribuído à classe (K) que obteve maior probabilidade. Os autores referem que este método é eficiente e robusto, com a vantagem de permitir a intervenção do utilizador na colocação e edição das sementes. O algoritmo apresenta uma melhor reprodutibilidade em estruturas bem definidas como a língua e pior desempenho em estruturas como os lábios. Na figura 3.9, observa-se, à direita e em cima, a segmentação 3D da língua obtida de forma manual (c) e, em baixo (f), pelo algoritmo semi-automático. Um exemplo da colocação das sementes, por um utilizador, numa imagem 2D *cine*, é apresentada em (d).

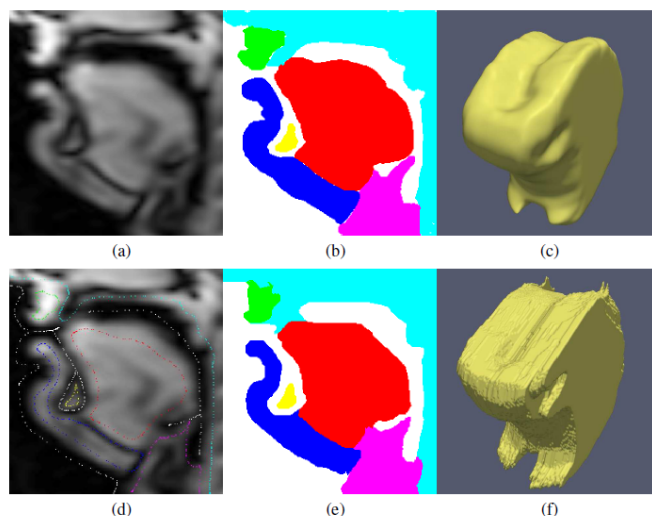


Figura 3.9: Segmentação com algoritmo de *Random Walker*. Visualização 3D da língua obtida a partir de segmentação manual e com o algoritmo semi-automático (em baixo, à direita). Fonte da imagem: Suo *et al.* (2011).

3.2.2 Estudos dinâmicos

A aquisição de imagens dinâmicas ou em tempo-real envolve um número muito grande de imagens a processar. Nesta situação, a segmentação efetuada manualmente é um processo muito laborioso e exige recursos humanos consideráveis (Narayanan *et al.*, 2004; Bresch *et al.*, 2008a). Todavia, alguns dos estudos consultados, recorreram a técnicas de delineação manual, por exemplo, para extração de contornos em imagens 2D, ao longo do tempo. Veja-se, por exemplo, Shadle *et al.* (1999), Delvaux *et al.* (2002a), Kedrova *et al.* (2009) e, mais recentemente, em Niebergall *et al.* (2012).

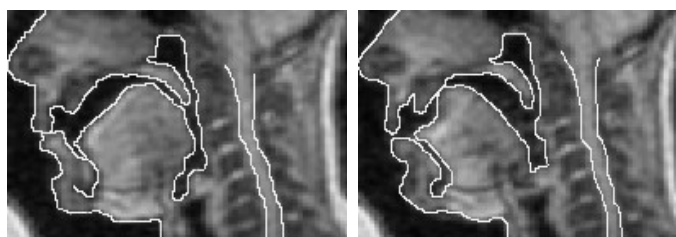


Figura 3.10: Exemplo de estudo em tempo real com segmentação manual do trato vocal. Fonte da imagem: Kedrova *et al.* (2009).

Por exemplo, no estudo de Kedrova *et al.* (2009), foram obtidas imagens a uma razão de 2.7 *frames/s*, num total de 500 imagens, relativas a vários fonemas orais e nasais da língua Russa. A figura 3.10 apresenta uma imagem obtida em tempo real em que o contorno foi delineado manualmente.

Em Niebergall *et al.* (2012), um estudo dirigido para o estudo de alguns fonemas da língua germânica, a resolução temporal é das melhores, considerando os estudos nesta área (cerca de 30 *frames* reconstruídas por segundo), todavia o processamento restringe-se à extração manual

de alguns contornos em imagens selecionadas. No fundo, apesar da excelente taxa de amostragem, não é explorado o potencial de informação dinâmica, contido nos dados adquiridos.

Engwall (2004), num estudo de modelação da língua, efetua a extração do contorno da língua, utilizando um algoritmo de deteção de fronteira, após *threshold* da imagem, seguido de correção manual.

Avila-García *et al.* (2004) efetuam a segmentação de imagens da língua, utilizando uma técnica que recorre a um algoritmo que conjuga *Active Shape Models* (ASM) e *Dynamic Hough Transform* (DHT). Para a criação do modelo da língua, foram utilizadas 39 configurações desta estrutura obtidos no âmbito de um estudo prévio. As segmentações da língua, que serviram para construir o modelo, foram obtidas de forma manual. Segundo os autores, a combinação de DHT com ASM permitiu melhorar os resultados da segmentação.

Vasconcelos *et al.* (2010) recorrem também a modelos estatísticos baseados em contornos ativos (ASM e AAM) para segmentar e extrair as características principais do trato vocal, num estudo com RM dinâmica. O modelo estatístico foi baseado em 23 configurações do trato vocal obtidas num falante e testado sobre imagens desse mesmo falante.

No que diz respeito ao processamento de imagens em tempo real, devemos destacar o trabalho do grupo da Universidade da Califórnia (SPAN) que tem feito uma aposta notável neste domínio. Assim, em Narayanan *et al.* (2004), o primeiro trabalho deste grupo que envolveu a aquisição de imagens RM em tempo real, é referida a importância da utilização de técnicas automatizadas para acompanhar as variações do trato vocal ao longo do tempo e identificar áreas ou regiões relevantes do trato, com interesse do ponto de vista linguístico e fonético. Nesse estudo, os autores recorrem a modelos deformáveis, num contexto de aplicação de filtros de *Kalman* e técnicas de *optical flow*, com o intuito de acompanhar e quantificar alguns eventos durante a articulação dos sons. Mais do que resultados fonéticos, apontam-se caminhos e possibilidades a explorar para obter aquisições de imagens de RM em tempo real e efetuar o seu processamento.

O método proposto em Narayanan *et al.* (2004) é depois aperfeiçoado e apresentado por Bresch *et al.* em 2006a. Resumidamente, após a iniciação manual efetuada na primeira imagem de determinada sequência, que permite definir os contornos de interesse, o algoritmo utiliza um procedimento de otimização que permite acompanhar os contornos de forma automática. Para realizar esta tarefa, o algoritmo utiliza informação a partir do gradiente de intensidade das imagens e do fluxo ótico (*optical flow*). Esta técnica permite estimar, *frame a frame*, o deslocamento de segmentos da imagem, através de um método de procura baseado na correlação entre imagens. Os parâmetros utilizados para otimização (informação relativa ao gradiente, *optical flow* e constrangimentos da curvatura) são manualmente definidos pelo operador. No caso do trato vocal, um dos maiores problemas com o algoritmo ocorre nas zonas de oclusão, ao longo do trato, uma vez que nestas zonas a intensidade do gradiente é zero (ou seja, não existe fronteira ar/tecido). Nesta situação, terá que existir correção por parte do operador, mas, segundo os autores, será sempre mais favorável do que uma segmentação puramente manual. Este algoritmo é depois utilizado no estudo de Lee *et al.* (2006), para o processamento de imagens, numa investigação cujo objetivo era avaliar o impacto das emoções (e.g., tristeza, felicidade) na articulação dos sons da

fala.

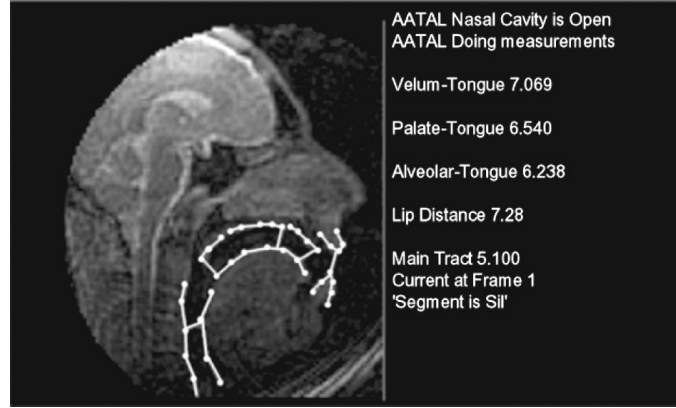


Figura 3.11: Segmentação de imagens 2D em tempo real com *kalman snakes* e *optical flow*. Fonte de imagem: Narayanan *et al.* (2004).

Em Bresch *et al.* (2008a) e Bresch e Narayanan (2009), é apresentada uma abordagem um pouco diferente das anteriores, com a implementação de uma técnica para detecção de contornos operando no domínio da frequência (*espaço K ou de Fourier*). A figura 3.12 ilustra os vários passos do processo. Para o aplicar ao estudo das vias aéreas, os autores tiveram que incorporar um modelo geométrico com informação anatômica para resolver o problema da otimização (*gradient anatomically informed gradient descent procedure*). A imagem da esquerda da figura 3.13 apresenta o modelo geométrico do trato vocal que referimos. À direita, são identificados os diferentes descritores do trato vocal numa aplicação posterior deste método (Ramanarayanan *et al.*, 2010).

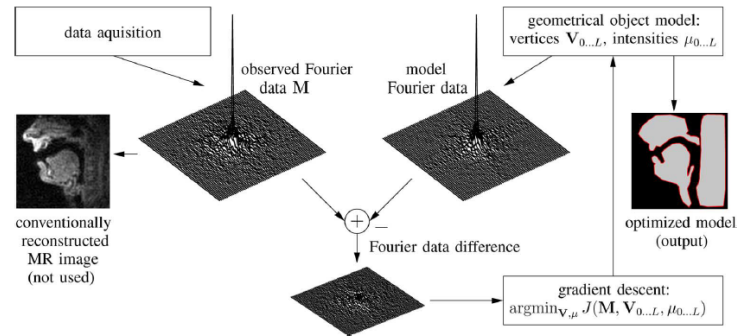


Figura 3.12: Segmentação no domínio da frequência. Fonte da imagem: Bresch e Narayanan (2009).

Esta técnica de segmentação tem sido utilizada em estudos posteriores da equipa, para processamento de imagens de RM em tempo real, para fins bastante diversificados como: avaliação do discurso emocional (veja-se, por exemplo, Lee e Narayanan, 2010); estudo de aspetos relacionados com posturas articulatórias (*articulatory settings*) como pausas (entre fala, intervalos entre segmentos não falados), postura de repouso, leitura e discurso espontâneo ou para comparar o planeamento e execução de discurso falado e cantado (Ramanarayanan *et al.*, 2010, 2011).

Partindo da premissa de que nem sempre será necessário extrair todo o contorno do trato vocal, para que se possam estudar alguns aspetos linguísticos e fonéticos relevantes (e.g., detetar

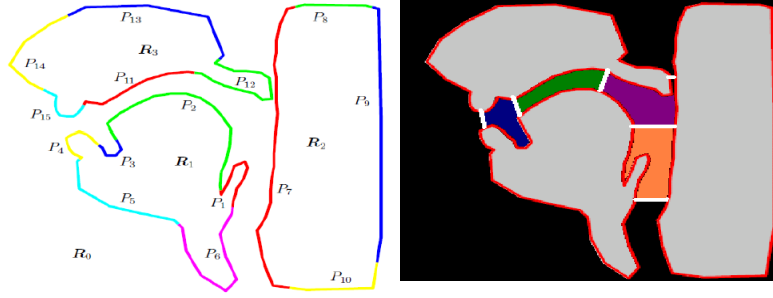


Figura 3.13: Modelo geométrico do trato vocal (esquerda) e descritores do trato vocal (direita). Fonte da imagem: Bresch e Narayanan (2009) e Ramanarayanan *et al.* (2010).

áreas de constrição ao longo do trato), foi proposto um método de processamento, baseado nas alterações de intensidade do *pixel* em regiões correlacionadas (*Data-driven analysis using correlated image regions*) (Lammert *et al.*, 2010b). Este método foi, mais tarde, aperfeiçoado por Proctor *et al.* (2011), permitindo também inferir acerca da trajetória dos articuladores e obter outro tipo de informação dinâmica. Estes métodos são considerados métodos simples, robustos e eficientes, permitindo encurtar o tempo dedicado ao processamento das imagens (quase automáticos). São suficientes para a obtenção e quantificação de informação relativa ao local e grau da constrição e de outros eventos articulatórios. Uma vez que permitem extrair dados relativos a diferentes zonas do trato (ou articuladores), em simultâneo, vêm possibilitar uma melhor compreensão dos gestos articulatórios e da coordenação entre estes.

O método apresentado em Proctor *et al.* (2011) foi aplicado por Hagedorn *et al.* (2011) no estudo de vários sons consonânticos do italiano (consoantes dorsais, alveolares, etc), num falante sem perturbações da fala e também em Hagedorn *et al.* (2012), neste caso num indivíduo com apraxia do discurso, estendendo o domínio de aplicação destas ferramentas ao estudo de perturbações articulatórias. A futura utilização destas ferramentas numa base de dados multi-modal (*real-time MRI, TIMIT*) recentemente recolhida (Narayanan *et al.*, 2011) é apontada por Proctor *et al.* (2011). A figura 3.14 ilustra a utilização do método, baseado na intensidade média do *pixel*, na determinação automática do local e grau de constrição, em consoantes do italiano.

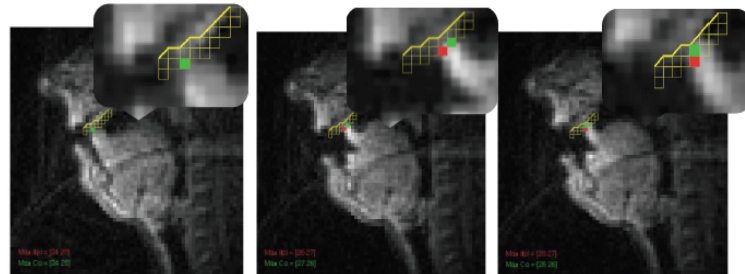


Figura 3.14: Identificação da zona e extensão da constrição na produção de sons consonânticos. O exemplo refere-se às consoantes coronais [t, n, l]. Fonte da imagem: Hagedorn *et al.* (2011).

Finalmente, uma referência ao estudo de Shosted *et al.* (2012b) que incluiu aquisição de

RM estática (3D) e dinâmica, para o Árabe. Os autores recorrem também a um método baseado na variação da intensidade média do *pixel* (*average pixel intensity*) ao longo do tempo, para processar as imagens em tempo real, com o intuito de obter informação relativa às articulações faríngeas, frequentes nessa língua. A aplicação de *Principal Component Analysis* (PCA) às variações de intensidade do *pixel*, ao longo do tempo, nas cinco regiões estudadas, demonstraram que 85 % da variação encontrada se fica a dever a duas regiões na faringe.

3.3 Os problemas de processamento de imagem nesta área

A segmentação do trato vocal e dos articuladores envolvidos durante a produção de fala constitui um grande desafio, quer seja na segmentação de imagens 2D e 3D, quer em imagens obtidas de forma dinâmica ou em tempo real (2D+t). Talvez, por essa razão, alguns estudos que foram referidos na secção anterior recorram a técnicas de segmentação manual ou não efetuem qualquer tipo de processamento sobre as imagens, como é o caso dos estudos de Demolin *et al.* (1997, 2000), Gregio (2006), Medeiros e Demolin (2006), ou Kedrova *et al.* (2006, 2008).

Interessa identificar as principais dificuldades e problemas, com que nos deparamos neste tipo de estudos, associados com o processamento das imagens adquiridas. Em primeiro lugar, os aspetos que têm a ver com a própria qualidade das imagens⁴ que é, usualmente, muito baixa, no que diz respeito a parâmetros ou atributos como a S/R e a resolução espacial (Narayanan *et al.*, 2004; Bresch *et al.*, 2008a; Bresch, 2011).

Poder-se-ia argumentar que é possível, no presente, obter volumes de dados RM com resolução isotrópica sub-milimétrica (e.g., Gold *et al.*, 2007), que nos forneceriam imagens com excelente resolução espacial e boa S/R, não constituindo, *a priori*, um grande problema para os atuais algoritmos de segmentação. No entanto, lembre-se que um dos grandes constrangimentos nos estudos de produção de fala é o tempo de aquisição, que terá que ser necessariamente baixo (da ordem dos *s* ou *ms*). Ora, como sabemos, devido a condicionantes de vária ordem tem que ser estabelecido um compromisso entre resolução temporal, espacial e a S/R. Assim, quando se pretende adquirir imagens com grande rapidez (elevada resolução temporal), terá que se sacrificar a resolução espacial (utilizando matrizes baixas, espessuras corte mais elevadas) ou fazer outras opções (e.g., menos amostras) que, invariavelmente, vão conduzir a uma diminuição da intensidade do sinal, com impacto na S/R. A resolução espacial e, particularmente o ruído são determinantes para a qualidade da imagem e têm uma grande influência no desempenho dos algoritmos de segmentação/delineação de objetos, em particular, os mais automatizados. A figura 3.15 permite ilustrar de que forma os compromissos estabelecidos entre cobertura anatómica, resolução temporal, espacial e S/R influenciam a qualidade das imagens. Por exemplo, a imagem adquirida com tempo de aquisição mais curto (44 *ms*) apresenta melhor resolução temporal, mas é óbvia a degradação da resolução espacial, presença de ruído e mais artefactos. Para além disso, a informação é

⁴A definição de qualidade de imagem e os conceitos inerentes à avaliação dessa qualidade, estão longe de serem conceitos absolutos e estão sempre ligados a questões de subjetividade. A qualidade de imagem é perçetualmente uma combinação ponderada de um número significativo de atributos (contraste, ruído,...) da imagem devendo ser tida em consideração o contexto ou aplicação em que foi adquirida (International Imaging Industry Association - IIA definition of image quality (tradução nossa)).

relativa a apenas um corte (reduzida cobertura anatômica).

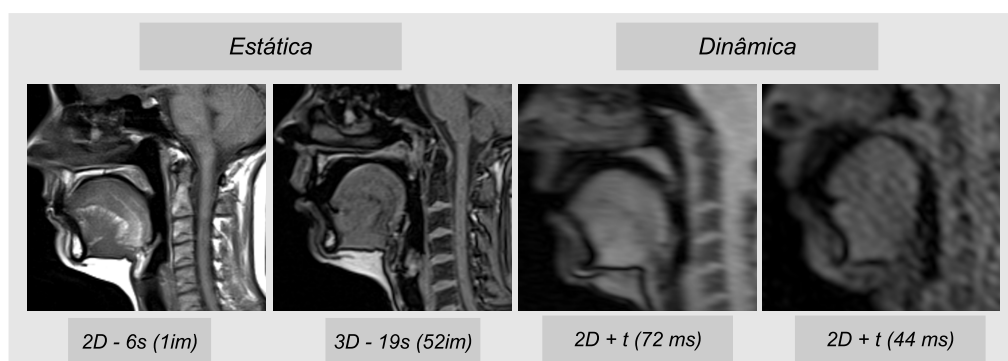


Figura 3.15: Exemplos de imagens, adquiridas no âmbito deste trabalho, apresentando informação obtida de forma estática (2D e 3D) e dinâmica (2D+t) durante a produção de fala.

Também os efeitos de volume parcial e a não-uniformidade da intensidade, característicos das imagens de RM, associados a múltiplos artefactos intrínsecos ao método, vêm complicar o processamento das imagens. Dos variados artefactos que podem surgir nas imagens (Zhuo e Gullapalli, 2006; Morelli *et al.*, 2011), de etiologia muito diversa (e.g., suscetibilidade magnética, desvio-químico, reconstrução, movimento), os de suscetibilidade magnética são particularmente graves na zona do véu palatino, uma vez que existe uma interface (ar/tecido) em que há grandes diferenças de suscetibilidade, podendo perturbar a visualização de um articulador (o véu palatino) que é da maior importância no estudo dos sons nasais. O próprio material utilizado nas reconstruções dentárias, presente na maior parte dos indivíduos, que usualmente não tem impacto nos estudos de RM, pode nestes estudos de produção de fala impedir (por severidade dos artefactos) uma boa visualização das estruturas dentro da cavidade oral (e.g., a língua). Acresce ainda a dificuldade em distinguir estruturas calcificadas (dentes) e o ar nas imagens de RM. Este facto pode não ser relevante para estudar algumas classes de sons, mas é extremamente importante em sons como as laterais, em que se estabelecem passagens laterais dos lados da língua - que interessa modelar, sendo fundamental identificar o que é ar e o que são dentes.

Para além de tudo isso, a própria geometria do trato vocal é muito complexa na produção de alguns sons, particularmente nas consoantes laterais e nos sons nasais. Na produção deste tipo de sons, formam-se descontinuidades ao longo do trato: nos sons nasais, porque o véu palatino está baixo, dando origem a duas cavidades separadas (oral e nasal); e nas laterais pela formação de passagens laterais separadas e cavidade supra-lingual. A figura 3.16 apresenta imagens obtidas durante a produção sustentada de fonemas. O volume de dados foi adquirido em 19 s. Pode evidenciar-se a dificuldade em diferenciar os dentes do ar, na cavidade oral (imagem no plano coronal, à esquerda) e a complexidade geométrica (e.g., descontinuidades) do trato vocal (imagem sagital, à direita).

Nas aquisições dinâmicas (e.g., tempo-real), não é necessário sustentar as articulações, mas existe movimento dos articuladores, podendo surgir artefactos devido a movimento quando a resolução temporal é insuficiente (considerando a velocidade dos articuladores). Por outro lado, as opções para acelerar as aquisições e as sequências de pulso seleccionadas podem dar origem a outro

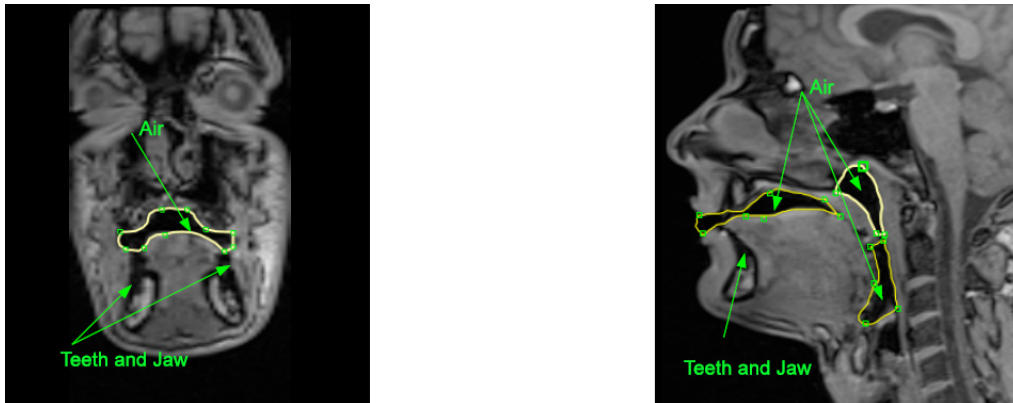


Figura 3.16: Exemplo ilustrativo de imagens do trato vocal adquiridas, neste estudo, durante a produção sustentada de fonemas.

tipo de artefactos que degradam a qualidade global da imagem, dificultando o seu processamento. Note-se que, nestas imagens, adquiridas de forma dinâmica é necessário acompanhar alterações de posição sofridas pelos vários articuladores ao longo do tempo, extrair contornos, detetar áreas de constricção e obter parâmetros relativos aos vários articuladores de forma sincronizada.

Finalmente, a quantidade de dados a processar, quer para aquisições 3D, quer em tempo real, é considerável (milhares de imagens). Nas aquisições volumétricas, o número de imagens por volume é elevado, uma vez que a área anatómica de interesse é extensa (necessário observar o trato vocal em toda a sua extensão, desde os lábios até à glote), usualmente são adquiridos vários sujeitos (para contemplar variabilidade entre sujeitos) e vários volumes de dados para cada sujeito (relativos a vários sons, contextos, posições) para viabilizar comparações.

Nas aquisições em tempo real, como é possível, na atualidade, obter taxas de amostragem elevadas, uns meros segundos de aquisição conduzem a um número significativamente elevado de imagens que têm que ser processadas. Por exemplo, no nosso estudo, a cada aquisição de 5 s, a uma razão de 14 *frames/s*, correspondem 75 imagens. As imagens da figura 3.17, obtidas no contexto das aquisições em tempo real, representam cerca de 1s de aquisição.

3.4 Métodos de segmentação e ferramentas utilizadas neste estudo

A escolha dos métodos utilizados para efetuar a segmentação da base de dados adquirida, no contexto do presente trabalho, contou muito com a experiência prévia adquirida no processamento deste tipo de imagens, na sequência do trabalho desenvolvido por Carbone (2008), que ocorreu paralelamente com o nosso trabalho de Mestrado (Martins, 2007). Os problemas e dificuldades que foram identificados na secção anterior resultam da nossa experiência com este tipo de imagens e vão ao encontro do que tem sido referido na literatura (e.g., Bresch *et al.*, 2008a).

Em primeiro lugar, devido à quantidade de dados a processar (vários falantes, sons e

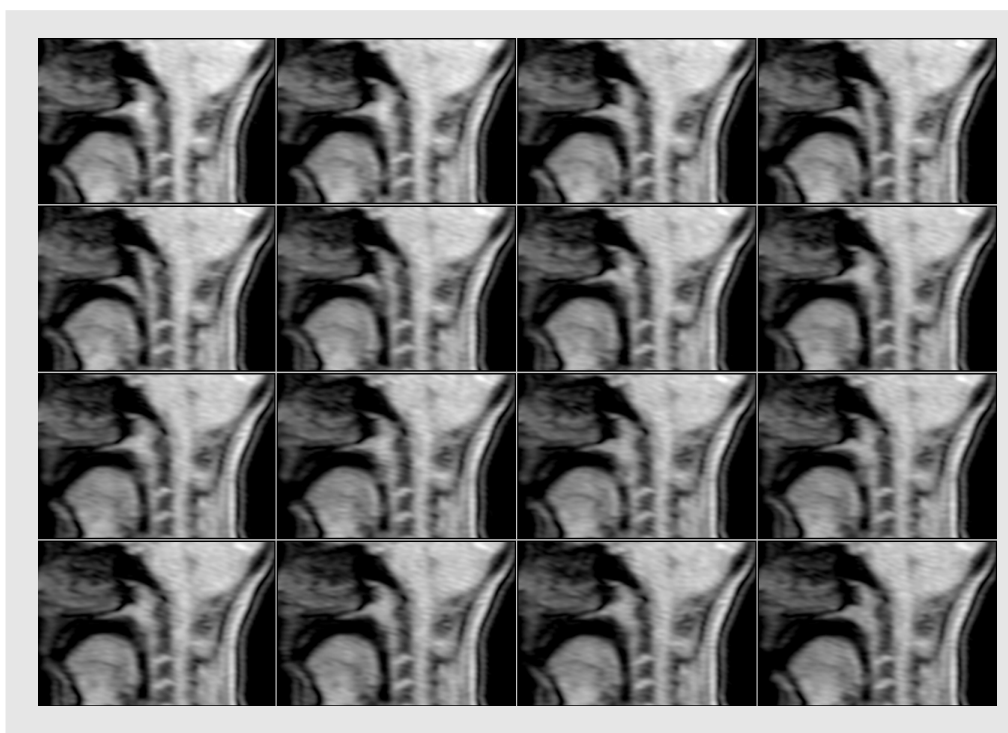


Figura 3.17: Exemplo de imagens adquiridas em tempo real, para um dos falantes (CO) deste estudo, durante a produção de fala com uma resolução temporal de 72 ms (14 frames/s). Imagens correspondentes a pouco mais que um segundo de aquisição.

contextos), a segmentação totalmente manual não se afigurou como hipótese a considerar. Assim, a opção pela exploração de vários métodos **semi-automáticos**, com maior ou menor nível de interação do utilizador, pareceu-nos ser a opção mais indicada para o tratamento desta base de dados. Contudo, como se verá mais adiante, nos capítulos 5 e 6, em algumas situações, as rotinas utilizadas revelaram-se muito pouco eficientes (e.g., extração e co-registo das arcadas dentárias dos falantes).

De entre os métodos semi-automáticos, os que se baseiam nos **modelos deformáveis** são os que, devido às suas características (veja-se McInerney e Terzopoulos, 1996, p.2), nos pareceram os mais ajustados para a segmentação de grande parte da base de dados adquirida. Como tivemos oportunidade de explicitar, lidamos com imagens de relativa baixa qualidade, provenientes de vários sujeitos, com estruturas anatómicas complexas e variadas, que variam muito na sua forma etc.

Considerando estes métodos, em função das características de cada um, efetuou-se uma seleção que ponderou as características das imagens (estáticas ou dinâmicas), as próprias estruturas a segmentar (e.g., língua ou trato vocal) e os objetivos que se pretendiam alcançar (e.g., áreas ao longo do trato).

Assim, dentro da classe dos modelos deformáveis, foram testados/utilizados neste trabalho algoritmos baseados em *Region Competition Snakes*, *Live-wire* e *Level-sets* para processamento dos volumes de dados. O *live-wire* tinha sido já utilizado, na nossa equipa, com bons resultados,

para a segmentação de volumes de qualidade semelhante, em particular, para obtenção de áreas ao longo do trato vocal (segmentação 2.5D). O recurso ao *live-wire 3D*, como se viu na secção 3.2.1, parece ser bastante promissor, para a segmentação dos volumes, podendo trazer uma melhoria de eficiência, quando comparado com a nossa abordagem. Para lá da maior eficiência, o facto de lidar bem com topologias complexas (como é o caso) torna-o um método apelativo. No entanto, seria necessário utilizar plataformas que incorporem essa rotina ou, em alternativa, proceder à sua implementação em MATLAB, o que seria mais difícil para quem, como a autora, não tem experiência de programação.

Relativamente às *Region Competition Snakes* e *level sets*, como se constatou, a sua utilização no âmbito de estudos de produção não é muito referida. No entanto, motivados pelos resultados que têm sido alcançados com estes métodos para modelação geométrica de várias estruturas anatómicas (e.g., Ribeiro *et al.*, 2009), considerámos que se seria uma alternativa a considerar.

Para além dos métodos baseados em modelos deformáveis, foi também utilizado o *Region Growing* (RG), fundamentalmente para a segmentação das imagens (2D) em tempo real. Neste caso, os resultados do estudo-piloto efetuado por Carbone (2008), com imagens semelhantes, demonstraram que esta rotina de segmentação, relativamente simples, poderia lidar bem com este tipo de imagens. O recurso a métodos mais automatizados seria o ideal (e.g., modelos estatísticos como ASM e AAM)⁵ e poderá ser uma alternativa a considerar para o processamento das imagens em tempo real. Contudo, será necessário utilizar informação proveniente de vários falantes, por forma a sustentar a criação de um modelo estatisticamente robusto que consiga lidar com a variedade anatómica e para as diferentes configurações do trato vocal na produção dos vários sons⁶.

3.4.1 ITK-SNAP e MeVisLab

Neste trabalho, o processamento das imagens foi efetuado recorrendo a duas ferramentas de processamento de imagem, o ITK-SNAP (versão 2.0.0)⁷ e o *MeVisLab* (versão 2.0)⁸, que têm implementados vários algoritmos para segmentação de imagem. Uma das razões para esta escolha resulta do facto de serem ferramentas de acesso livre. As versões utilizadas estão já relativamente estáveis, havendo, no entanto, módulos que apresentam alguns problemas (e.g., 3D *live-wire* no *MeVis*). As duas ferramentas são versáteis, apresentam interfaces gráficas (Graphical User Interface - GUI) intuitivas e relativamente fáceis de utilizar por indivíduos que não têm conhecimentos alargados na área da programação, como é o caso da autora deste trabalho. Apresentam também uma curva de aprendizagem relativamente rápida, o que se revelou muito importante. Apesar de poderem ser utilizadas noutras áreas, o domínio de aplicação destas ferramentas é, essencialmente, a imagem médica. Apresentamos uma breve descrição de cada uma delas.

⁵Estes métodos apesar de terem um grau de automatismo superior aos tradicionais modelos deformáveis são, mesmo assim, dependentes da qualidade do treino e da inicialização. Note-se que o treino assenta normalmente na marcação de *landmarks* por vários observadores, preferencialmente, experientes na área.

⁶Na realidade, este trabalho já foi iniciado por Silva e Teixeira (2013) a partir das imagens adquiridas, em tempo real, no contexto deste trabalho.

⁷Disponível para *download* em <http://www.itksnap.org>.

⁸Disponível para *download* em <http://www.mevislab.de>.

O *MeVisLab* foi desenvolvido, em conjunto, pelo *Fraunhofer MeVis* e *MeVis Medical Solutions AG*, em Bremen, na Alemanha. O *MeVisLab* é uma ferramenta poderosa, que pretende contribuir para o desenvolvimento de algoritmos de processamento de imagem, visualização e métodos de interação, com enfoque na imagem médica e, em particular, na área da radiologia. O *MeVis* assenta numa estrutura modular, suportando vários níveis de programação: 1) para definição de interfaces com o utilizador, recorre a uma linguagem de *scripting* própria (MDL); 2) para controlo dinâmico dos vários módulos e até algum processamento primário, o ambiente *MeVisLab* incorpora a linguagem *Python*; e 3) para o desenvolvimento de novos módulos de processamento, pode ser utilizada linguagem *C++* e a própria biblioteca de processamento de imagem que integra o pacote *MeVisLab*.

Para além dos módulos básicos de processamento e visualização, tem integrados módulos mais avançados de segmentação, registo, avaliação quantitativa e análise de imagem funcional. Suporta imagens *Digital Imaging and Communications in Medicine* (DICOM) e permite a integração no *Picture Archiving and Communication System* (PACS). Uma das suas componentes mais importantes é a plataforma *MeVis Image Processing Library* (ML), que tem disponíveis mais de 300 módulos, permitindo a utilização de vários filtros de imagem (e.g., filtragem linear e não linear, filtros morfológicos), várias técnicas de segmentação (e.g., *region growing*, *live-wire*, *fuzzy connectedness*), transformações (e.g., *Affine*, *Radon*, *manual registration*), entre outras possibilidades. Utilizando a sua interface, é possível conjugar vários dos módulos disponíveis, criando redes personalizadas pelo utilizador. Apresenta um elevado desempenho em bases de dados de grande dimensão (informação disponível em Manual *MeVisLab*, 2010).

A Figura 3.18 ilustra a interface gráfica do *MeVis*, com uma rede criada com o intuito de efetuar uma reconstrução multiplanar (MPR) curva, no contexto do nosso trabalho.

O *ITK-SNAP* (*Snake Automated Partitioning*) é uma aplicação baseada na utilização de contornos ativos. O *SNAP* foi desenvolvido com o intuito de poder constituir uma alternativa viável à segmentação manual, particularmente na segmentação de estruturas anatómicas, em contexto clínico. As primeiras aplicações foram no âmbito da neuro-imagem (e.g., segmentação do corpo caloso e ventrículos laterais). Depois disso, tem sido muito utilizado, com sucesso, como ferramenta de segmentação na modelação geométrica de estruturas das mais variadas regiões anatómicas como coluna vertebral, ossos longos, mandíbula, entre outras (Lopes *et al.*, 2007; Fernandes, 2008; Ribeiro *et al.*, 2009). Os testes de validação já efetuados evidenciaram uma elevada precisão nas segmentações, excedendo mesmo a segmentação manual em termos de eficiência e reprodutibilidade (Yushkevich *et al.*, 2006).

O *SNAP* implementa dois métodos de segmentação 3D, que já aqui tivemos oportunidade de referir: 1) os contornos ativos geodésicos ou modelos geométricos deformáveis, propostos por Caselles *et al.* (1997) e 2) as *Region Competition Snakes*, apresentadas por Zhu e Yuille (1996).

A evolução do contorno, em ambos os métodos é governada pela ação de forças externas e internas, no entanto os dois métodos diferem na forma como definem estas forças. Os contornos ativos geodésicos funcionam bem, quando as estruturas apresentam um contorno com intensidade de sinal elevada (*edge-based*). Por outro lado, as *region competition snakes* funcionam melhor

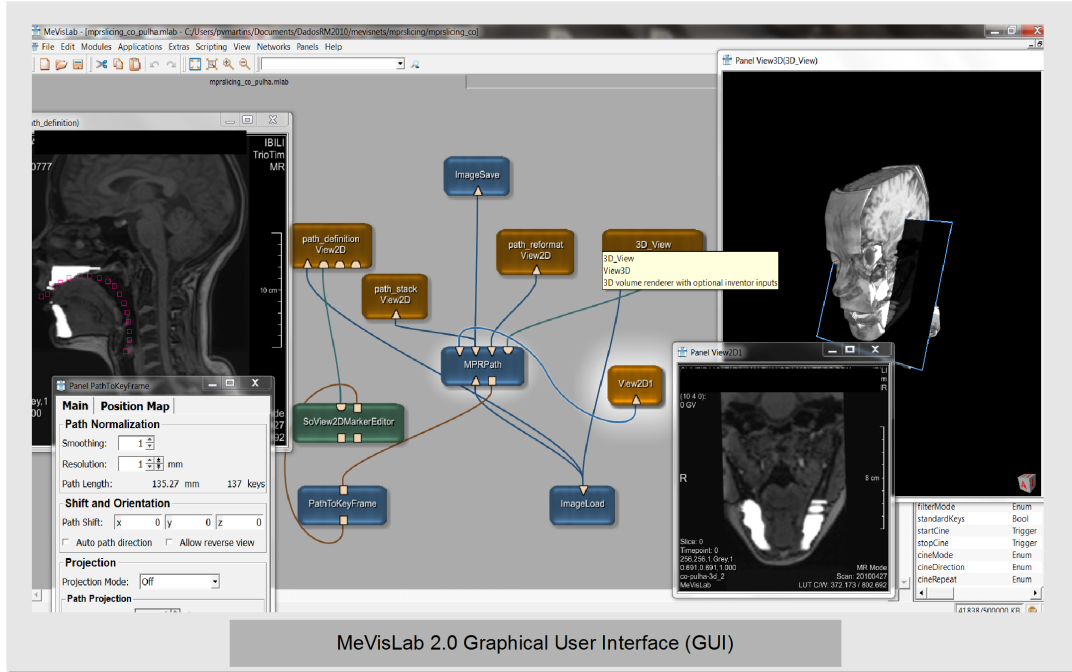


Figura 3.18: Interface gráfica (GUI) do *MeVisLab*. Rede desenvolvida para efetuar MPR curvo. Exemplo efetuado para um dos falantes participantes neste estudo, relativo ao som [L].

quando a estrutura a segmentar apresenta uma gama de intensidades bem definida, em relação ao fundo (Yushkevich *et al.*, 2006). Um exemplo de segmentação do lobo caudado, utilizando as duas abordagens é apresentado na figura 3.19.

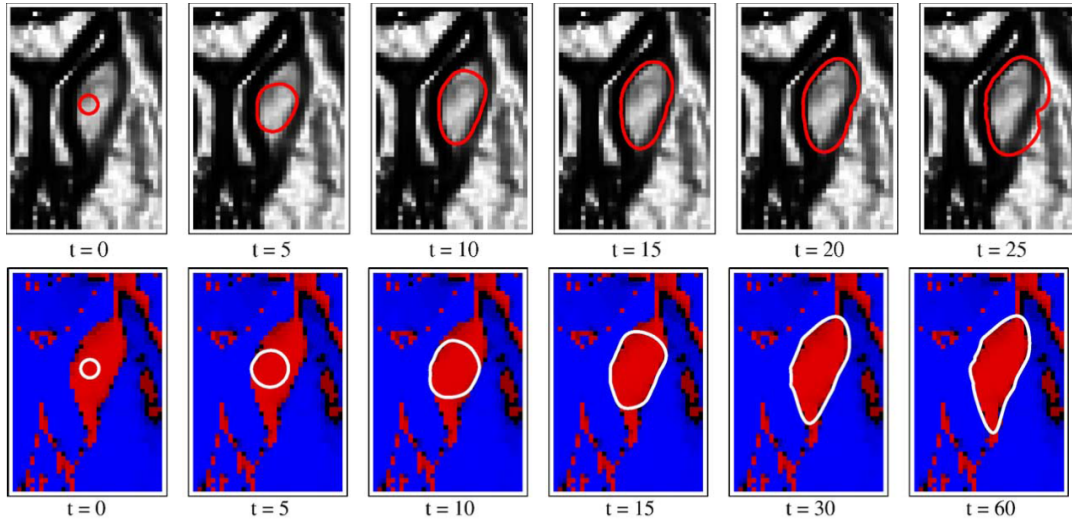


Figura 3.19: Exemplo de segmentação do lobo caudado com os dois algoritmos disponibilizados pelo *ITK-SNAP*. Fonte da imagem: Yushkevich *et al.* (2006).

Mais adiante, no capítulo 5, será efetuada uma descrição mais detalhada relativa à utilização destas duas ferramentas, no contexto deste trabalho, para a segmentação das várias estruturas de interesse (e.g., língua, trato vocal).

Em suma, o processamento de imagem é um passo fundamental para extrair informação, qualitativa e quantitativa, a partir das imagens obtidas com RM, no contexto da produção de fala. O processamento destas imagens é complexo, devido a questões relacionadas com a própria qualidade das imagens (largamente condicionada pelos tempos de aquisição curtos), pela complexa geometria do trato vocal, em particular para algumas classes de sons e pelo elevado volume de dados que é, usualmente, recolhido. Nesta área, o recurso à segmentação manual tem sido frequente e a utilização de métodos baseados em modelos deformáveis é bastante mais recente e começa agora a ter mais expressão. Os métodos baseados em modelos deformáveis são os que, devido às suas características, melhor se poderão ajustar e que podem oferecer uma solução mais satisfatória. A melhor qualidade das imagens e a utilização de novas abordagens (e.g., aquisições 3D) fazem com que estes métodos tenham um maior potencial de aplicação nesta área. O pressuposto de que nem toda a informação contida nas imagens necessita ser explorada tem conduzido a novas abordagens (*data-driven*) que facilitam o processamento e a extração da informação.

Capítulo 4

Consoantes Laterais e Vogais Nasais

“After you’ve used a term like tongue height – raise the tongue, lower the tongue, move it to the back or front – you begin to feel that is what your tongue is actually doing. But you are kidding yourself. This is not really what one does in trying to produce a vowel...”

PETER LADEFOGED, 1985 (INTERVIEW)

No presente capítulo, efetuar-se-á uma revisão acerca do conhecimento atual relativo à produção das consoantes laterais e vogais nasais, com base em dados para outras línguas e, sempre que existam, para o Português Europeu (PE). Será conferida uma especial atenção à descrição articulatória destes sons, complementando-se a informação, sempre que necessário, com uma breve referência às suas propriedades acústicas.

Os sons da fala podem ser divididos em vogais e consoantes. As vogais são produzidas sem que existam obstáculos significativos, ao longo do trato vocal, à passagem do fluxo de ar proveniente dos pulmões. Na produção dos sons consonânticos, por outro lado, existe sempre uma constricção que condiciona o fluxo de ar na cavidade faríngea ou na cavidade oral (Andrade e Viana, 1996; Moutinho, 2000; Mateus *et al.*, 2005).

Do inventário de sons do PE, constam nove vogais orais¹ ([e] **pêga**, [E] **leva**, [i] **pipo**, [o] **tôpo**, [O] **pote**, [u] **luva**, [a] **pato**, [ɔ] **cada**, [ɪ] **devi**) e cinco vogais nasais ([ɐ̃] **campa**, [ẽ] **pente**, [ĩ] **sinto**, [õ] **ontem** e [ũ] **mundo**). Em relação às consoantes, o PE possui três consoantes nasais ([m] **cama**, [n] **cana**, e [ɲ] **banho**), seis consoantes fricativas ([f] **faca**, [v] **vaca**, [s] **sala**, [z] **zarpa**, [ʃ] **chá** e [ʒ] **janela**), seis consoantes oclusivas ([p] **pato**, [b] **bola**, [t] **taco**, [d] **dado**, [k] **caro**, [g] **gato**), duas consoantes laterais ([l] **laca** e [ʎ] **palha**) e, também, os sons róticos ou vibrantes ([ʀ] **cara** e [ʀ̥] **carro**). Podemos ainda referir as semivogais ou glides (sons articulatoriamente semelhantes às vogais, mas que não podem constituir núcleo de sílaba, nunca ocorrem sozinhas e

¹Neste trabalho, optou-se por utilizar, ao invés do Alfabético Fonético Internacional - AFI, o alfabeto fonético **X-SAMPA** que é uma versão estendida do SAMPA (*Speech Assessment Methods Phonetic Alphabet*) (Wells, 1995), disponível em <http://www.phon.ucl.ac.uk/resource/phonetics/>. Este alfabeto utiliza caracteres do código ASCII não sendo necessário a utilização de fontes específicas.

não podem receber acento) que se associam às vogais orais ou nasais para formar ditongos orais e nasais, bastante abundantes na língua portuguesa (Moutinho, 2000; Mateus *et al.*, 2005).

Apesar das semelhanças com outras Línguas Românicas, existem algumas características específicas do Português que interessa salientar:

“The characteristics which at first hearing distinguish the pronunciation of Portuguese from that of the other Western Romantic languages [are]: (a) the very large number of diphthongs (...); (b) the large number of nasal vowels and nasal diphthongs; (c) frequent alveolar and palatal fricatives (...); and (d) the extremely dark quality of the common variety of l-sound” (Stevens, 1954, p.6)

No entanto, alguns aspetos relativos a estas características tão específicas e importantes da língua Portuguesa não foram ainda totalmente estudadas ou alvo de investigações que recorram a técnicas instrumentais e que forneçam informação para lá das descrições impressionistas.

Como já foi referido, no capítulo introdutório deste trabalho, não será possível dedicar atenção a todos os sons, abordando todas as questões que mereceriam interesse ou necessitariam de clarificação. Desta forma, por razões distintas, o nosso interesse recaiu sobre as consoantes laterais e as vogais nasais.

No caso das consoantes laterais, a efetiva escassez de dados de produção para o Português motivou e fundamentou a nossa escolha. Mesmo para outras línguas, estes sons não têm sido muito estudados, sobretudo tendo em conta a quantidade de estudos existentes para outras consoantes ou para as vogais (Cagliari, 1977a; Zhou, 2009; Recasens, 2012). Para além disso, estudar sons como as laterais constitui um grande desafio, uma vez que: 1) estas consoantes apresentam configurações articulatórias e características acústicas mais complexas (e.g., presença de anti-formantes) do que a maior parte dos outros sons (Narayanan *et al.*, 1997a; Zhou, 2009); 2) estão associadas a uma grande variabilidade articulatória (Zhou, 2009)² e acústica (Andrade, 1999; Recasens, 2012) motivadas pela sílaba, contexto, acento, falante, entre outros (Andrade, 1997); 3) a par com as vibrantes³, são, também, os sons que as crianças adquirem mais tardiamente, que implicam maiores dificuldades de articulação e os mais afetados quando existem distúrbios fonológicos (Wertzner *et al.*, 2007; Guerreiro e Frota, 2010; Lousada *et al.*, 2012).

No que diz respeito às vogais nasais, não é a escassez de dados que nos move, uma vez que já existem variados estudos sobre o tema para o PE. Neste caso, o que se pretende é complementar a informação já recolhida com EMA, mediante a aquisição de dados com RM, para mais falantes e focalizando-se na obtenção de informação acerca da dinâmica dos articuladores. Como já tivemos oportunidade de referir, a articulografia não nos permite a observação de todo o trato vocal, sendo a informação restrita a poucos pontos nos lábios, língua e véu palatino. Desta

²De acordo com Zhou (2009), esta variabilidade tem um grande impacto, por exemplo, no contexto da síntese articulatória, uma vez que das várias configurações possíveis, para um mesmo som, alguma poderá dar origem a um som simulado mais natural e de melhor qualidade.

³As consoantes laterais /l/ e /L/ e algumas vibrantes como o [4] são frequentemente agrupadas numa classe designada como consoantes líquidas. Todavia, apesar de frequente nas referências consultadas, essa designação não é consensual.

forma o comportamento dos articuladores, no seu todo, é inferido e não diretamente observado e medido.

4.1 As consoantes laterais

As laterais são sons caracterizados pela presença de uma constrição ao longo da linha médio-sagital, com a formação de passagens laterais de um ou dos dois lados do dorso da língua. Para a maior parte das línguas, a oclusão verifica-se na região dento-alveolar e é produzida pelo contacto entre a ponta da língua (articulação apical) ou da lâmina (articulação laminal) e os dentes ou alvéolos (Cagliari, 1977a; Narayanan *et al.*, 1997a; Stone e Lundberg, 1996; Andrade, 1999; Ladefoged, 2001; Zhou, 2009). Para além deste tipo de articulação, muitos outros pontos de articulação são referidos como possíveis para a produção das consoantes laterais: região pós-alveolar, palatal e mesmo na região velar (Ladefoged e Maddieson, 1986).

No PE, do ponto de vista fonológico, existem duas consoantes laterais: a lateral palatal /L/, produzida através de uma oclusão central entre a parte anterior do dorso da língua e o palato anterior; e a lateral alveolar /l/, que apresenta uma constrição na parte anterior do trato vocal, produzida pela aproximação da ponta da língua à região dento-alveolar.

4.1.1 A lateral alveolar /l/

4.1.1.1 Características articulatórias

Grande parte das descrições encontradas na literatura para a lateral alveolar são baseadas em informação 2D (e.g., traçados radiográficos ou de RM 2D), relativa a pontos (e.g., XRMB ou EMA) ou restrita à cavidade oral (e.g., EPG). No entanto, para alguns autores, quando se trata de uma lateral a informação 2D não será suficiente para uma descrição detalhada desta consoante, uma vez que grande parte dos acontecimentos articulatorios ocorre para além da linha média (Narayanan *et al.*, 1997a; Badin *et al.*, 2002, 2006; Zhou, 2009).

Do ponto de vista da configuração articulatória, para além do ponto e extensão do contacto, importará considerar: 1) o número de passagens ou canais laterais que se formam e a sua extensão; 2) a configuração da língua imediatamente atrás da zona de contacto e 3) a configuração geral da língua, não só em 2D, mas idealmente em 3D (Stone e Lundberg, 1996; Narayanan *et al.*, 1997a; Zhou, 2009). De facto, já Cagliari (1977a) referia que uma descrição detalhada da articulação das laterais deveria contar com informação relativa à contração dos músculos da língua, particularmente dos bordos em relação ao centro e com as variações ao nível da cavidade faríngea. Segundo alguns autores, a caracterização do /l/ passará, ainda, por considerar aspetos temporais (Sproat e Fujimura, 1993; Browman e Goldstein, 1995) e efeitos contextuais (Recasens, 2004; Recasens e Espinosa, 2005; Recasens, 2012).

São descritas, tradicionalmente, duas variantes da lateral alveolar: uma não velarizada

(*clear* ou *light*, no inglês), que ocorre em início (ou ataque) de sílaba (e.g., *lata*), e outra velarizada⁴ (*dark*), produzida em final (ou coda) de sílaba (e.g., *canal*).

Como foi já referido, a produção do /l/ está sujeita a uma grande variabilidade, motivada, essencialmente, pela posição que a consoante ocupa na sílaba. Apesar da variabilidade e das diferenças articulatórias entre as duas realizações do /l/, há alguns aspetos comuns às duas produções (Narayanan *et al.*, 1997a), nomeadamente: 1) existência de contacto alveolar; 2) forma convexa da parte média e posterior do dorso da língua e 3) compressão lateral em direção à linha média (estreitamento da língua). A questão do estreitamento da língua (na zona da lâmina) na produção das laterais, para que seja permitido fluxo de ar lateral, já tinha sido, também, apontada por Sproat e Fujimura (1993). A própria definição de consoante lateral apresentada por Ladefoged e Maddieson (1986) sugere a existência desta compressão da língua:

“... *laterals are sounds in which the tongue is contracted in such a way as to narrow its profile from side to side so that a greater volume of air flows around one or both sides than over the center of the tongue.*” (Ladefoged e Maddieson, 1986, p.95)

Stone e Lundberg (1996) referem que a forma da língua na produção da lateral alveolar é única, comparada com as outras consoantes, uma vez que apresenta elevação dos segmentos anterior e posterior, com um pequeno *groove*, indicando compressão do segmento médio. No entanto, Narayanan *et al.* (1997a) afirmam que aspetos como o *flattening* ou *grooving* da língua, posteriores à zona do contacto, variam muito entre os sujeitos, sendo, portanto, características secundárias dependentes da atividade muscular da língua ou do tipo de articulação produzida (mais apical ou laminal).

Configuração do dorso lingual atrás do contacto

A principal diferença entre as duas variantes do /l/ reside, essencialmente, na configuração da língua atrás do contacto alveolar (Giles e Moll, 1975; Sproat e Fujimura, 1993; Narayanan *et al.*, 1997a). O /l/ velarizado pode ser produzido com diferentes graus de abaixamento do pré-dorso da língua e retração ou elevação da parte posterior do dorso em direção à faringe ou à região uvular, dando origem a duas cavidades separadas atrás da constrição (cavidade oral e cavidade faríngea). No /l/ não velarizado, o dorso da língua está mais anteriorizado e elevado, formando uma única cavidade posterior à região de contacto alveolar (Recasens, 2004, 2012).

Recasens (2012) refere que a configuração articulatória da variante velarizada é semelhante à de um /u/ e a não velarizada semelhante a um /i/. A figura 4.1 apresenta a configuração articulatória, para as duas variantes do /l/.

⁴De acordo com Recasens *et al.* (1996), a utilização dos termos velarizado e *dark* não deveria ser efetuada de forma indiferenciada como se verifica na maior parte das vezes. Segundo este autor o termo velarização implica uma elevação ativa do dorso da língua em direção ao véu palatino. Na produção de um *dark-l* existe sempre um movimento de retração do dorso da língua, mas a direção dessa retração não é sempre em direção ao véu palatino podendo ser em direção à faringe. Assim, quando se fala das características articulatórias do /l/ deveriam utilizar-se os termos velarizado e faringealizado e reservar-se o termo *dark* para fazer referência à qualidade da lateral.

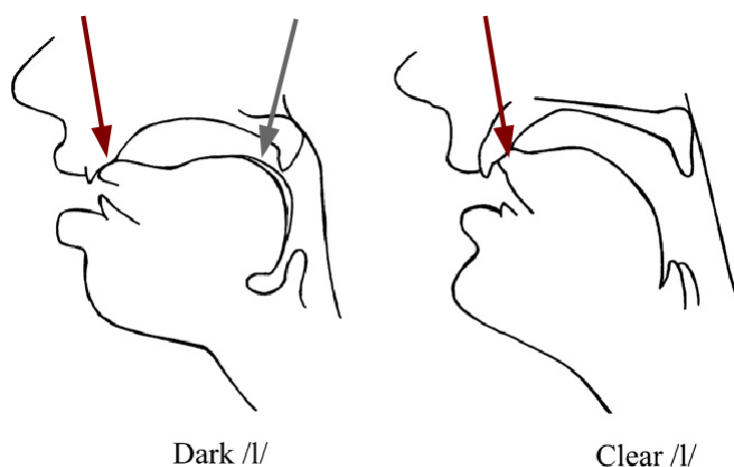


Figura 4.1: Configurações do trato vocal para o /l/: velarizado (*dark-l*) - à esquerda, e não velarizado (*clear-l*) - à direita. Fonte da imagem: Recasens (2012). As setas (acrescentadas à imagem) assinalam as zonas de constrição.

Ponto de articulação

A constrição anterior, para as duas variantes do /l/, ocorre, geralmente, na região dento-alveolar, podendo a articulação ser apical e laminal. Narayanan *et al.* (1997a) verificaram, para quatro falantes do Inglês Americano, que durante a produção do /l/ houve contacto da ponta da língua ou da lâmina lingual com a região alveolar ao longo do plano médio-sagital. Estes autores referem que a apicalidade ou laminalidade estará mais relacionada com as características dos falantes do que com a posição silábica ou com a velarização. Neste mesmo estudo, verificou-se, contudo, que a extensão e a área do contacto foi maior para as articulações laminais do que para as apicais. No caso das articulações laminais, a zona de contacto estende-se lateralmente, através da região médio-palatal.

Zhou (2009), para um falante do Inglês Americano, verificou que o /l/ velarizado foi produzido com uma articulação apical, enquanto o /l/ não velarizado foi articulado com a lâmina da língua. Para além disso, a extensão do contacto, na linha média e nos lados da língua, foi maior para o /l/ não velarizado, à semelhança do verificado por Narayanan *et al.* (1997a).

Recasens (2004) e Recasens e Espinosa (2005) salientam que articulações mais anteriores (dentais) poderão favorecer a velarização ou faringealização. Os dados coligidos pelos autores utilizando EPG sugerem que a constrição alveolar para o *dark-l* (e.g., Inglês Americano de Nova Iorque) é mais anterior, menos extensa e menos variável do que o observado para o /l/ não velarizado como se verifica, por exemplo, no Italiano.

O facto do contacto alveolar poder ser mais curto e menos efetivo na produção do /l/ pós-vocálico em conjunto com um movimento mais lento da língua poderá conduzir a situações em que há perda de contacto alveolar. Este fenómeno é designado por redução articulatória (*articulatory undershoot*) ou (semi)vocalização. O processo de redução articulatória no /l/ pós-vocálico parece corroborar a teoria que considera o gesto de velarização do /l/ como antagonista dos movimentos

de anteriorização e da elevação da ponta da língua (Leidner (1976) referido por Carter, 2002).

Giles e Moll (1975), num estudo cineradiográfico, constataram que, frequentemente, na produção do /l/ pós-vocálico não existiu contacto alveolar, sobretudo em taxas de elocução mais elevadas e em contexto de vogais baixas.

No estudo de Narayanan *et al.* (1997a), apesar dos contactos serem menos extensos para o *dark-l*, apenas em um dos quatro falantes não se verificou contacto linguo-alveolar. O facto de, neste estudo, o processo de vocalização não ter sido frequente, estará, na interpretação de Wrench e Scobbie (2003), relacionado com o método de aquisição, já que as produções foram sustidas.

A vocalização pode também ser entendida como uma perda do traço consonântico na variante *dark* do /l/ (Nakamura, 2009; Garcia, 2010). Considerando que o gesto da ponta da língua é consonântico e o gesto de recuo (não chega a existir constrição) do dorso é vocálico, quando a ponta da língua não atinge o seu alvo (*target*) restará um gesto vocálico.

A vocalização do /l/, em posição de coda, é bastante frequente no Português do Brasil (Cristófaros-Silva e Oliveira, 2001; Hora, 2006; Hahn e Quednau, 2007; Pinho e Margotti, 2010), com exceção de alguns dialetos do Sul (Pinho e Margotti, 2010) onde a variante velarizada é dominante.

Áreas ao longo do trato vocal e passagens laterais

Considerando agora informação 3D, em particular as áreas transversais ao longo do trato vocal durante a produção do /l/, Narayanan *et al.* (1997a) afirmam que parece existir um padrão nas funções de área, comum para a maioria dos falantes (3 dos 4 participantes), com uma diminuição das áreas na zona uvular e porção superior da faringe e aumento das áreas na zona palatal. Para o /l/ produzido em posição final de sílaba (*dark-l*), as áreas na região palatal (imediatamente atrás do contacto alveolar) são mais elevadas do que para o *light-l* devido a uma maior compressão da língua e menor extensão dos contactos laterais com o palato. Também, as áreas na região uvular ou zona superior da faringe são significativamente mais pequenas para o *dark-l* quando comparadas com as observadas no *light-l*, devido a uma maior elevação e/ou recuo da língua na variante *dark*. Esta diminuição das áreas reflete-se também ao longo da zona velar ou inferior da faringe dependendo da parte da língua envolvida na formação da constrição.

Relativamente às passagens laterais, estas são observadas nas duas variantes do /l/. Segundo Narayanan *et al.* (1997a), as cavidades são assimétricas quer em comprimento quer em área. As áreas destas variam entre 0.1 a 0.5 cm², dependendo da morfologia dos sujeitos e do tipo de articulação. Desta forma, como as articulações laminais foram, em geral, mais extensas, também as passagens laterais, nesta situação, apresentaram maior extensão.

Note-se, no entanto, que neste estudo, apenas foram medidas as passagens laterais que se formam de cada um dos lados da língua, na região do contacto alveolar. Os autores referem, todavia, que estas passagens laterais se prolongam para trás do contacto alveolar convergindo com o fluxo de ar central na região palatal. A extensão das passagens laterais é condicionada pela dimensão dos contactos laterais da língua com o palato, que foram mais extensas na variante não

velarizada do que na velarizada.

Uma modelação mais completa das passagens laterais foi conduzida por Zhou (2009), com o objetivo de observar os efeitos da dimensão e área destas, bem como da cavidade supra-lingual, no espectro acústico. As medidas foram obtidas a partir de dados RM, considerando as posições do /l/ em ataque e em coda. O autor verificou que as passagens laterais no *dark-l* eram mais curtas, restringindo-se ao local do contacto alveolar (com aproximadamente 1cm). Neste caso, como não existe contacto lateral da língua com o palato, não há formação de cavidade supra-lingual. No caso do *light-l*, as passagens laterais eram mais compridas e, devido aos contactos da língua com o palato, lateralmente, forma-se uma cavidade supra-lingual, dando origem a três cavidades separadas. À semelhança do estudo de Narayanan *et al.* (1997a), Zhou (2009) verificou que as passagens laterais são, geralmente, assimétricas em área e extensão, mas maiores no *light* do que no *dark-l*. A questão do impacto das passagens laterais no espectro acústico do /l/ será abordada na secção 4.1.1.2.

Efeitos coarticulatórios

Um outro aspeto que tem sido referido como diferenciador entre as duas variantes do /l/ diz respeito à sua maior ou menor resistência a efeitos coarticulatórios. Uma das primeiras referências relativas à variação do /l/ em função do contexto vocálico em que é produzido foi efetuada por Bladon e Al-Bamerni (1976). Estes autores mostraram que o *dark-l* é mais resistente aos efeitos da vogal adjacente do que o *light-l*, tendo por base as diferenças de F2 do /l/ produzido em contexto das vogais /i/ e /a/⁵.

Recasens *et al.* (1996) argumentam que, no caso do /l/ velarizado, devido à existência de uma constrição alveolar e de recuo ou elevação da língua (constrição dorsal), o dorso da língua ficará de tal forma estrangido que o torna altamente resistente aos efeitos da vogal, em particular quando a vogal adjacente é o /i/, uma vez que os gestos para produzir os dois segmentos são antagónicos.

No *dark-l*, existe menos variação no local da constrição, acontecendo invariavelmente o contacto na zona dento-alveolar no contexto das vogais /i, a, u/ (e.g., Inglês Americano), enquanto que no *light-l* (e.g., Espanhol) há mais variação com a vogal adjacente (Recasens e Espinosa, 2005). Estes autores compararam, também, dois dialetos do Catalão (um *dark* e outro *light*) e verificaram que na variante *dark* (Catalão de Maiorca) o /l/ foi resistente à coarticulação com a vogal (tendo em consideração a localização da constrição, o padrão de contacto dorso palatal e os valores de F1 e F2), quer em posição inicial quer em posição final. Na variante *light* (Catalão de Valência), o /l/ apresentou efeitos vocálicos em qualquer das posições silábicas consideradas. Segundo os autores, estes resultados corroboram a hipótese que nas línguas ou dialetos que apresentem um /l/ fortemente velarizado não existem dois alofones posicionais. Mais recentemente, como resultado de um estudo alargado que envolveu dados acústicos recolhidos para diferentes línguas e dialetos,

⁵Bladon e Al-Bamerni (1976) introduziram o conceito de resistência à coarticulação e propuseram um índice (MCD - *Mean Coarticulatory Distance*) que estima o grau de coarticulação. Este é baseado nas diferenças de F2 para o /l/ em contexto de /i/ e /a/. Este índice tem sido, posteriormente, utilizado por Recasens em vários estudos de coarticulação que envolvem a lateral alveolar (Recasens, 2004; Recasens e Espinosa, 2005).

com variantes *dark* e *light* do /l/, Recasens (2012) verifica que há menos variabilidade de F2, menos diferenças posicionais e menos efeitos coarticulatórios na variante *dark* do /l/. Os efeitos da vogal adjacente sobre o /l/ diminuem na seguinte progressão: posição intervocálica > posição inicial > posição final.

Aspetos temporais e de coordenação entre os gestos

Para além das diferenças articulatórias, reportadas entre as variantes do /l/, têm merecido grande destaque na literatura os aspetos temporais e de coordenação entre os gestos envolvidos na produção da lateral alveolar (coordenação intergestual). De facto, alguns autores defendem que mais do que as diferenças na amplitude dos gestos, são os diferentes padrões de coordenação temporal, verificados para as duas variantes do /l/, que podem constituir a sua diferenciação (Sproat e Fujimura, 1993; Browman e Goldstein, 1995; Gick, 2003).

Sproat e Fujimura (1993) referem que na produção do /l/, independentemente da posição silábica, estão envolvidos dois gestos: um do ápice da língua (o gesto apical) e outro relativo à retração e abaixamento da parte anterior do dorso da língua, designado por gesto dorsal. Assim, no caso do /l/ produzido em posição final, o movimento do dorso (gesto vocálico) precede o movimento de elevação da ponta da língua (gesto consonântico) enquanto que, em posição inicial, os gestos ocorrem em simultâneo ou o gesto da ponta da língua precede ligeiramente o gesto do dorso (Browman e Goldstein, 1995; Sproat e Fujimura, 1993; Gick, 2003).

Gick *et al.* (2006), numa investigação com consoantes líquidas que incluiu várias línguas e dialetos - Inglês do Oeste Canadano (IOC), Francês do Quebec (FQ), Coreano, Servo-croata (SC), Mandarin de Beijing e Squamish Salish (SS)⁶ - com o objetivo de avaliar os padrões de organização gestual em função da filiação silábica, constataram que, no geral, todas as línguas apresentaram dois gestos na produção do /l/, com pequenas diferenças. No entanto, no que diz respeito às questões temporais não se verificou a existência de um padrão generalizável que se aplique a todas as línguas.

Archangeli e Berry (2010) referem que o contraste *dark/light* nas laterais poderá ser conseguido de várias formas. O *light-l* poderá não incluir um gesto dorsal e, em algumas línguas, o padrão de coordenação diferenciado entre ataques e codas, relevante para o inglês, poderá não se verificar (e.g., Georgiano).

A este respeito, Recasens (2012) afirma que serão ainda necessários mais estudos para verificar até que ponto a implementação temporal do *light* e *dark-l* varia nas diferentes línguas e dialetos, particularmente se a antecipação do gesto dorsal em relação ao gesto apical verificado para o /l/ em coda (*dark-l*) dependerá ou não do grau de velarização da consoante.

⁶O SS é uma das cerca de 23 línguas Salishan que são ou foram faladas pelos povos indígenas na zona do Pacífico Noroeste dos Estados Unidos. O SS é ainda falado pelo povo Squamish que habita na zona sudoeste de British Columbia e a oeste de Vancouver (Kroeber, 1999).

A velarização como um *continuum*

Até este momento, efetuou-se uma comparação entre as variantes do /l/, assumindo que o /l/ em início de sílaba é geralmente não velarizado (*light*) e que em final de sílaba é velarizado (*dark*). No entanto, esta visão dicotômica, que implica a existência de dois alofones categoricamente diferentes, não é partilhada por todos os autores (e.g., Sproat e Fujimura, 1993). Estes autores defendem que existe uma única entidade fonológica que é foneticamente implementada como mais *light* ou mais *dark* em função de fatores como a posição na sílaba, a duração e o contexto prosódico em que o /l/ é produzido, podendo existir um *continuum* de possíveis pronúncias do /l/ dentro da mesma língua.

Para além disso, Recasens (2004), Recasens e Espinosa (2005) e Recasens (2012) sugerem que também entre as várias línguas/dialetos existem realizações intermédias: “... *that darkness ought to be considered a gradual phonetic property, rather than a categorical attribute. According to this view, there is a dialect-specific continuum proceeding from strongly dark to very clear varieties of /l/ through consonant realization exhibiting intermediate degrees of darkness...*” (Recasens e Espinosa, 2005, p.2).

Com base numa série de dados disponíveis para diferentes línguas e dialetos, Recasens e Espinosa (2005) consideram três grandes grupos de línguas, tendo em atenção as realizações do /l/ nas diferentes posições silábicas.

1. Um grupo de línguas que apresenta um /l/ fortemente velarizado independentemente da posição silábica. Este é o caso do Escocês, Inglês Americano, Russo, Polaco, Albanês e Búlgaro. Destes, alguns apresentam basicamente a mesma realização em posição inicial e final de sílaba (e.g., Russo, Inglês de Leeds e Português Europeu), outros (como o Inglês Americano) apresentam uma realização ligeiramente mais *clear* em posição inicial do que em posição final;
2. Dialetos em que a realização do /l/ é diferente, em função da sua posição na palavra. Neste grupo incluem-se várias variedades do Inglês Britânico⁷ (Inglês da zona de Newcastle e do Sul de Inglaterra) e, também, o Catalão da zona oriental;
3. Um terceiro grupo de línguas e dialetos que apresentam um /l/ não velarizado com a mesma realização ou, muito semelhante, do /l/ em posição inicial e em posição final de sílaba. É o caso do Alemão, Espanhol, Francês, Italiano, entre outros.

4.1.1.2 Características acústicas

A lateral alveolar é caracterizada acusticamente por apresentar uma primeira formante (F1) baixa, mas com valores ainda assim mais elevados do que os correspondentes a uma vogal alta; valores

⁷No que diz respeito ao Inglês Britânico a existência de alguma variedade na produção do /l/, já tinha sido apontada por Wells (1982) que refere zonas do país em que o /l/ apresenta realizações marcadamente *light* em todas as posições silábicas e outras claramente *dark* em todas as posições. Ainda, segundo o mesmo autor, para o Inglês da Escócia o /l/ é, geralmente, velarizado em todos os contextos prosódicos não existindo a alternância *light-dark* observada na pronúncia representativa do Inglês.

de F2 não muito afastados dos de F1, e um valor de F3 que é mais elevado do que na maioria das vogais (Espy-Wilson, 1992; Prahler, 1998; Zhou, 2009).

Relativamente às duas variantes do /l/, os valores de F1 e F3 são mais elevados para o /l/ não velarizado do que para a sua variante velarizada. O F2 é mais baixo para o /l/ velarizado, uma vez que F2 está diretamente relacionado com a posição da língua na direção antero-posterior (mais ou menos recuada). De acordo com Stevens (1998) a existência de um ramo lateral (devido às passagens laterais) dará origem a um zero por volta dos 2200-4400 *Hz*, sendo o valor exato dependente do comprimento desse ramo lateral.

O impacto das passagens laterais e da cavidade supra-lingual no espetro acústico tem sido alvo de análise, sobretudo através de estudos de modelação acústico-articulatória, em especial para o Inglês Americano (Bangayan *et al.*, 1996; Prahler, 1998; Zhang *et al.*, 2003; Zhou, 2009; Zhou *et al.*, 2010). Zhou (2009) procurou modelar as variantes *light* e *dark* do /l/, considerando o número de passagens laterais, a sua extensão e assimetria e os efeitos da formação de uma cavidade supra-lingual separada. Este autor mostra que canais laterais com comprimento de 1 a 2 *cm* não produzem nenhum zero na região de F3-F5 para a lateral alveolar. Para que surja um zero nesta gama de frequências, os canais laterais terão que ser assimétricos, com um comprimento que poderá variar entre os 3-6 *cm* tendo que existir, para além disso, uma constrição estreita ou um fechamento completo no trato vocal.

No entanto, para Browman e Goldstein (1995), a existência de uma ou de duas passagens laterais será meramente uma consequência do estreitamento da língua para conseguir a constrição alveolar e o recuo do dorso. Ou seja, estando presentes estes dois gestos, o /l/ pode ser perceptualmente percebido sem a adição de passagens laterais. Os autores basearam-se em simulações informais efetuadas com uma versão modificada do modelo do trato vocal dos laboratórios Haskins - HLsyn, um modelo proposto, originalmente, por Mermelstein (1973).

Uma primeira abordagem de síntese da lateral alveolar para o PE foi efetuada, com recurso ao SAPWindows e no HLsyn, após definição dos gestos articulatórios, tendo por base o modelo (TAsk Dynamics Application) - TADA (veja-se Oliveira, 2009). Devido às limitações deste modelo, mais especificamente a impossibilidade de controlar a variável “forma de constrição” e, portanto, dificuldade em simular as passagens laterais, os resultados do teste perceptual efetuado não foram muito animadores. Serão necessários mais dados articulatórios para o PE, de modo a rever a configuração proposta, para além de desenvolvimentos nos próprios sintetizadores para acomodar a síntese destes sons.

4.1.1.3 O /l/ do Português Europeu

No Português Europeu, a lateral alveolar /l/ é descrita como sendo articulada na linha média com contacto do ápice ou lâmina lingual na região alveolar. No entanto, existe alguma variabilidade no que respeita à definição do ponto de articulação, de acordo com o referido em Oliveira (2009).

Também em relação à variação da lateral alveolar em função da sua afiliação silábica as posições são controversas. Assim, e de acordo com algumas descrições para o PE (Barbosa,

1994; Cunha e Cintra, 1997; Mateus e d'Andrade, 2000; Mateus *et al.*, 2005), são considerados dois alofones para o /l/: um não-velarizado que ocorre em início de sílaba ou ataque (e.g., como na palavra **laca**) e, outro, velarizado em coda ou posição final de sílaba (e.g., como na palavra **mal**). Na produção deste /l/ velarizado existe uma elevação da parte posterior do dorso da língua em direção ao véu palatino (Mateus *et al.*, 2005).

Barbosa (1994) afirma (referindo-se ao movimento do ápice em relação aos alvéolos no /l/ em posição final):“... *quando a língua toma a mesma direção mas o ápice não chega a tocar em qualquer ponto e a sua massa recua na cavidade bucal, obtém-se uma lateral velarizada...*”. Em posição intervocálica, a lateral será, também, velarizada, mas menos do que em posição final. A descrição de Barbosa parece indicar que, no /l/ em posição final, o gesto apical é reduzido.

No entanto, descrições mais antigas do PE consideram que uma das características distintivas do português é a existência de um /l/ fortemente velarizado “... *the extremely dark quality of the commoner variety of l-sound* ...” (Stevens, 1954 p.6). Mais recentemente, também Emiliano (2009) assume que o /l/ no PE será sempre velarizado, seja produzido em coda ou em ataque. Contudo, estas descrições são essencialmente impressionistas (baseadas nas capacidades auditivas dos investigadores) e não ancoradas em dados obtidos com recurso a técnicas instrumentais.

Os resultados do estudo acústico efetuado por Andrade (1999), com base nos dados de 3 informantes de Lisboa, apontam para a hipótese de o /l/ ser velarizado não só em coda, mas também em posição de ataque. Com base nestes dados, tendo em atenção os valores de F2 reportados, Recasens e Espinosa (2005) colocam o PE no grupo de línguas em que o /l/ é fortemente velarizado nas duas posições silábicas, conforme já referido na secção 4.1.1.1.

Mais recentemente, os dados articulatórios obtidos através da utilização de RM, para um falante (Martins, 2007; Martins *et al.*, 2008), e para dois falantes (Oliveira, 2009) apontam para a existência de uma única realização velarizada do /l/. A posição recuada e elevada da língua em direção ao véu palatino (ver figura 4.2) é observada nas imagens 2D sagitais obtidas para o /l/, quer em ataque (palavra **laço**), quer em coda (palavra **mal**). Também as funções de área (informação 3D) para o /l/, produzido nas duas posições, não revelaram diferenças em função da posição, sendo que em ambas existe uma diminuição das áreas na região velofaríngea, em consequência da elevação ou recuo do dorso da língua em direção à região velar (Martins *et al.*, 2008; Oliveira, 2009). O facto de, em contexto VCV⁸, se ter verificado grande resistência a efeitos contextuais (particularmente num dos falantes), também aponta para a hipótese de a lateral alveolar ser velarizada. No entanto, em qualquer dos estudos o número de falantes era pequeno, os *corpora* não foram desenhados com o intuito de estudar as consoantes laterais e a forma como as palavras foram utilizadas para elicitacão dos sons não permitiram efetuar generalizações.

Num estudo com EMA, especificamente concebido para o estudo da lateral alveolar, em que foram contemplados vários contextos e posições silábicas, para dois falantes do PE, os resultados não foram conclusivos, uma vez que cada um dos falantes apresentou um padrão de coordenação temporal diferente (Oliveira *et al.*, 2010).

Os resultados de dois estudos acústicos muito recentes (Marques, 2010; Monteiro, 2012),

⁸Vogal-Consoante-Vogal.

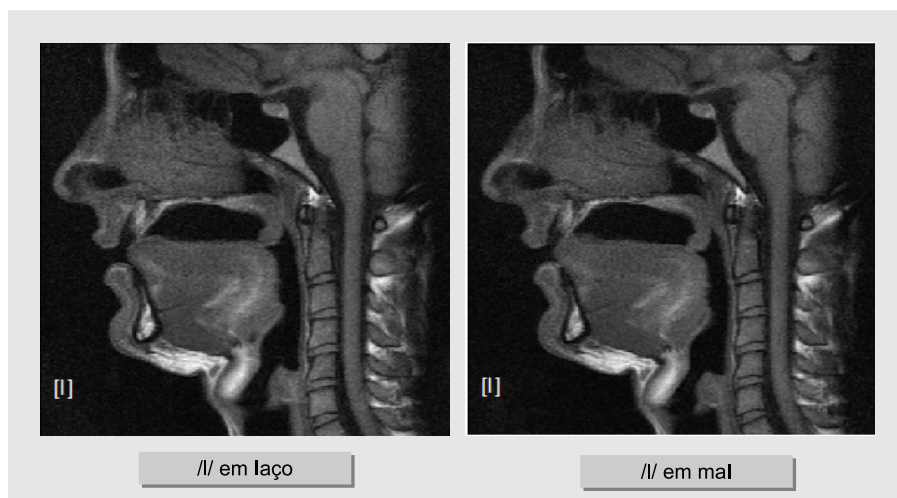


Figura 4.2: Configuração articulatória 2D do /l/ em posição inicial e posição final para um falante do PE. Fonte da imagem: Martins (2007).

que decorreram em paralelo com este trabalho, apontam para a hipótese da lateral alveolar ser velarizada, independentemente da posição silábica, considerando os valores de F2 (por volta dos 1000 Hz), obtidos nas duas investigações.

4.1.2 A lateral palatal - /L/

As consoantes laterais palatais fizeram parte do inventário da maior parte das línguas românicas. Atualmente, em algumas línguas (como o Francês) estas consoantes praticamente desapareceram e noutras (e.g., alguns dialetos do Espanhol) tendem a ser substituídas por uma glide palatal ou por uma fricativa palatal. Para o Português Brasileiro, Português Europeu e para o Catalão, estas consoantes têm sido mantidas em maior ou menor grau (Colantoni, 2004)⁹. A explicação deste fenómeno é complexa e vai para além do que são os objetivos deste trabalho, pelo que nos focalizaremos na sua descrição articulatória, fazendo breve alusão às suas características acústicas.

4.1.2.1 Características articulatórias

Ao contrário da lateral alveolar, que é objeto de interesse de uma quantidade de investigações, muitas das quais utilizam RM, para a lateral palatal não encontramos descrições articulatórias, baseadas em RM, que incluam este som, com exceção de estudos de carácter mais generalista, como o apresentado por Martins *et al.* (2008) para o PE, ou por Romano e Badin (2009) para o

⁹Para o caso concreto do PB, existem uma série de investigações recentes focalizadas no estudo da lateral palatal, mas direcionadas, em particular, para as suas possíveis realizações fonéticas, considerando vários dialetos, aspetos socioculturais e geolinguísticos. Assim, de acordo com as descrições, o segmento /L/ pode ser realizado como lateral palatal [L], como semivogal ou glide [j] - designado como processo de iotização, como lateral alveolar [l], como lateral alveolar palatalizada [lj] ou ainda como zero fonético (∅) - o que corresponde ao seu apagamento, entre outras possibilidades. Apesar da existência de todas estas variantes, cuja frequência de ocorrência é variável com os dialetos considerados, com a escolaridade e contexto socioeconómico, a lateral palatal [L] tem sido mantida no PB e continua a ser a variante mais produzida. É considerada a variante de prestígio, ocorrendo preferencialmente em zonas citadinas, grupos socialmente mais favorecidos e associando-se a um nível de escolaridade mais elevado (Brandão, 2007; Freire, 2011; Stein, 2011; Santos e Chaves, 2012).

Italiano. Também é certo que a maior parte dos estudos efetuados com os sons laterais são para a língua inglesa e do inventário fonético desta língua não fazem parte as consoantes laterais palatais (Stein, 2011).

Assim, para esta revisão recorreremos, fundamentalmente, a estudos baseados em imagens radiográficas (e.g., Ladefoged, 1988), palatografia (e.g., Cagliari, 1977a), EPG (e.g., Reis e Espesser, 2006; Recasens, 1993) e estudos acústicos (e.g., Silva, 1996) encontrados para o Português Brasileiro, Espanhol (incluindo alguns dialetos), Italiano e Catalão que, tal como o PE, preservam em maior ou menor grau este som no seu inventário.

A lateral /L/ é tradicionalmente incluída numa classe mais abrangente de sons, designada como consoantes palatais. Desta classe de sons fazem parte para além da lateral, segmentos nasais (como a consoante [J]), consoantes orais [C], a glide [j] e, ainda, outros sons, em função das diferentes línguas.

A descrição das características articulatórias das consoantes palatais não parece ser consensual entre os foneticistas como é afirmado em Recasens (1990), Recasens *et al.* (1993) e reforçado por Colantoni (2004). Um dos pontos de divergência está relacionado com o ponto de articulação destes segmentos e, o outro, com o facto de alguns autores os incluírem num classe de sons complexos (com dupla articulação), opinião defendida por Keating (1988), enquanto outros os consideram sons simples (Recasens, 1990; Recasens *et al.*, 1993; Recasens e Espinosa, 2006).

Relativamente ao ponto de articulação, em Ladefoged (1988) e Ladefoged (2001) considera-se que todos estes sons apresentam um ponto de articulação comum: o palato duro. No entanto, Recasens (1990; 1993; 2006) tem vindo a argumentar que é necessário uma maior especificação relativamente ao ponto de articulação das palatais, defendendo que estas deveriam ser divididas, pelo menos, em dois grupos: um grupo de consoantes alvéolo-palatais onde caberiam a consoante nasal [J] e a lateral [L] e outro, designado por “palatais verdadeiras” (*true palatals*) ou “palatais propriamente ditas” (*palatals proper*) onde se inclui, por exemplo, a glide [j].

Quando comparado com o [J], a articulação do [L] é mais anterior (Recasens e Espinosa, 2006). Esta diferença estará relacionada, segundo a mesma fonte, com constrangimentos no modo de articulação: uma articulação mais anterior facilitará o abaixamento dos lados da língua (a nível do pré-dorso) e, desta maneira, permitirá a passagem do fluxo de ar lateralmente, através dos canais laterais. Recasens e Espinosa (2006) consideram que esta tendência para uma articulação mais anterior, não se verifica só para o Catalão, mas também para outras línguas Românicas referindo-se, por exemplo, às descrições de Cagliari para o PB.

Para Cagliari (1977a), a lateral palatal é produzida com um movimento da parte anterior da língua em direção à zona alvéolo-palatal, estando a ponta da língua abaixada e mantendo pressão contra os incisivos inferiores. O facto de a articulação do [L] ser tendencialmente mais anterior, poderá tornar a lateral palatal mais semelhante a uma lateral alveolar palatalizada e esse facto poderá, em parte, explicar porque é que, por vezes, alguns falantes do PB a produzem como [l], particularmente no contexto da vogal [i]. Cagliari (1977a) chega a referir que a maior parte das descrições existentes para a lateral palatal, para outras línguas, não explicam o padrão por si encontrado, em palatogramas, para esta variante da língua portuguesa.

A extensão e área do contacto para o [L] é considerada grande se comparada com a lateral alveolar. Em relação ao [J], a extensão do contacto do dorso da língua no [L], particularmente, nas zonas média e pós palatal é menor do que o observado para a nasal. Este facto deve-se, mais uma vez, à necessidade de existir no [L] espaço para a passagem do ar, lateralmente, ao longo do trato vocal (Recasens, 1990; Recasens *et al.*, 1993).

Segundo Cagliari (1977a) e Recasens (1990), este “escape” do ar pode ser efetuado através de uma passagem lateral, delimitada internamente pelo bordo externo dos dentes e externamente pelas paredes internas da bochecha mas também, apesar de não ser uma característica comum em todas as línguas, escapando por cima dos molares, quando a língua não entra em contacto com essas estruturas. Cagliari (1977a) explica como se efetua a passagem do ar na produção de uma lateral palatal da seguinte forma:

“... whilst the tip of the tongue remains lowered, maintaining pressure against the lower incisors, the front of the tongue rises against the alveolo-palatal zone. The borders of the tongue make contacts along the back teeth and their inner alveolar edges on both sides, so that the airstream is not allowed to pass out over the borders of the tongue and out over the back teeth. The back of the tongue and the inner back walls of the mouth do not make contact with each other. On the contrary, there is a muscular action that produces a slight cheek swelling. As the airstream pressure increases in the mouth, the air start to escape laterally round the back teeth on both sides, through the channel formed by the external surface of the teeth and the inner walls of the cheek, and out of the mouth.” (Cagliari, 1977a, p.114)

Reis e Espesser (2006), utilizando EPG, prestando homenagem ao trabalho pioneiro de Cagliari, verificaram que o /L/, em falantes do PB, é um fone pós alveolar, com padrão de contacto transversal na zona alveolar e pós alveolar mas que não se estende por toda a zona de articulação (ver figura 4.3). Lateralmente, o contacto é irregular e mais longo, estendendo-se até à zona velar. Ao contrário da lateral alveolar, em que a zona de contacto é pequena e há a passagem do ar pelos bordos laterais da língua na zona dorsal, na palatal a obstrução estende-se mais posteriormente deixando um espaço reduzido para passagem do ar.

De forma semelhante à lateral alveolar, a lateral palatal é considerada uma consoante que apresenta alguma resistência a efeitos contextuais (Recasens e Espinosa, 2009; Martins *et al.*, 2011b).

4.1.2.2 Características acústicas

No que concerne à descrição acústica desta consoante, a informação também não é abundante. Esta escassez é referida por Colantoni (2004), adiantando como razão o facto de serem sons pouco comuns nas várias línguas. De acordo com os dados disponíveis para o PB, Espanhol e Italiano, esta autora refere que uma característica comum observada nas três línguas, para esta consoante, é a sua duração, consideravelmente mais longa que a da lateral alveolar e das consoantes líquidas em geral.

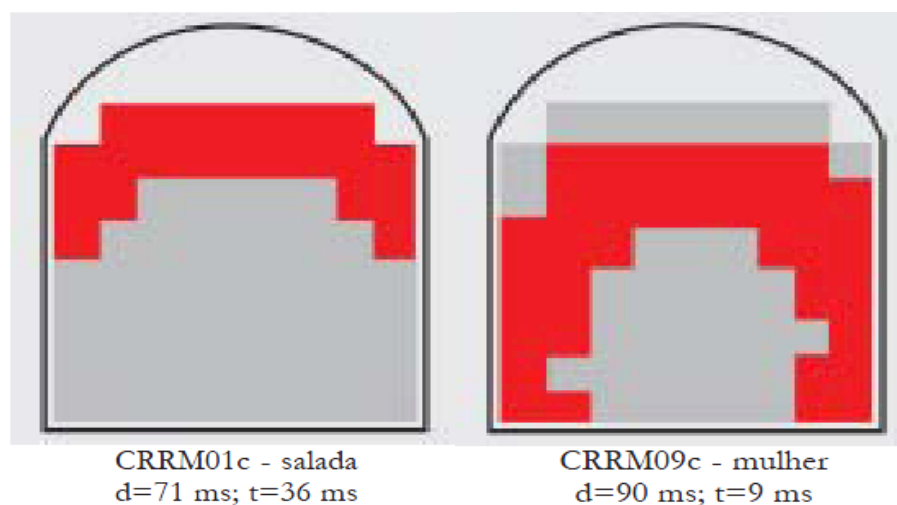


Figura 4.3: Padrão de contacto da lateral alveolar (à esquerda) e da lateral palatal (à direita) num falante do Português Brasileiro. Fonte da imagem: Reis e Espesser (2006).

Os valores de F1 situaram-se, para as referidas línguas, abaixo ou próximo dos 300 *Hz*; os de F2 variaram entre os 1870 *Hz* para o PB, 2000 *Hz* para o Espanhol e foram um pouco mais elevados para o italiano (entre 2273-2666 *Hz*); os valores de F3 foram, na generalidade, um pouco mais elevados do que os de F2 variando entre 2540 a 2766 *Hz* para o Espanhol e 2847 *Hz* para o PB. Os valores indicados para o Espanhol e Italiano são os referidos por Colantoni (2004), tendo por referência os trabalhos de Quilis et. al (1979) para o Espanhol e Bandon e Carbonaro (1978) para o Italiano. Os valores relativos ao PB foram consultados em Colantoni (2004) e confirmados em Silva (1996).

Rogers (2006), num estudo que visava estudar as características acústicas das consoantes líquidas, para o Wanyi (dialeto do Norte da Austrália), refere que a consoante lateral palatal apresenta um padrão acústico semelhante ao de uma vogal alta e anterior como o /i/, com valores de F1 baixos (abaixo dos 400 *Hz*), F2 e F3 relativamente próximos, em valor (acima dos 2000 *Hz*) e em amplitude. Distingue-se da lateral alveolar pelos valores de F1 mais baixos, valores mais elevados de F2 e F3 e pelas transições entre formantes.

4.1.2.3 O /L/ no Português Europeu

Para o PE, como já referimos anteriormente, a informação relativa a esta consoante é muito pouca. As descrições articulatórias que existem são as que geralmente são encontradas em obras genéricas de fonética. No que respeita a consoantes palatais, para além da lateral, existe para o PE a nasal palatal [J]. A lateral /L/, devido a impedimentos de natureza fonotáctica¹⁰, ocorre maioritariamente em posição intervocálica como nas palavras *palha*, *abelha*, ou *pulha*. Surge em posição inicial num número limitado de morfemas lexicais, que ocorrem raramente, e nos pronomes pessoais *lhe* e *lhes* (Viana, 1883; Mateus e d'Andrade, 2000). Segundo Viana (1883),

¹⁰Para o PB, a lateral palatal também é das consoantes menos produtivas, estando sujeita a constrangimentos fonotáticos de forma semelhante ao que se passa no PE (Brandão, 2007; Santos e Chaves, 2012).

pode ainda ocorrer em posição inicial em palavras importadas do Espanhol, como por exemplo em **lhano** (sortudo) ou **lhama**. Não se observa em posição final uma vez que, para o PE, apenas três consoantes se podem realizar nesta posição, o [4] como em **mar**, a lateral alveolar [l] (e.g., **mal**) e a fricativa [ʃ] como na palavra **más** (Mateus e d'Andrade, 2000).

Articulatoriamente, ela é tradicionalmente, descrita como uma consoante dorsopalatal ou simplificada como palatal (Sá-Nogueira, 1938; Barbosa, 1994; Cruz-Ferreira, 1999; Moutinho, 2000; Mateus *et al.*, 2005; Emiliano, 2009). Sá-Nogueira (1938), baseado em palatogramas, efetua a descrição da lateral dorsopalatal oral da seguinte forma:

“O ápice da língua apoia-se nos incisivos inferiores; o pré-dorso da língua adapta-se à região pré-palatal, para impelir a saída de ar segundo a linha média do canal bucal; os bordos laterais da língua abaixam-se para permitirem a saída de ar por ambos os lados ou só por um ... a língua no seu conjunto toma, por antecipação, a posição preparatória para a produção do fonema que se lhe segue...” (Sá-Nogueira, 1938, p. 55)

Com base em traçados médio-sagitais provenientes de imagens radiográficas, Barbosa (1994) descreve assim a lateral:

*“No mesmo ponto onde se produz a oclusiva [J], produz-se também a **contínua dorsopalatal**, não nasal (...) durante cuja realização o ar passa pelos dois lados, ou por um só, da língua. É pois como [l] uma **lateral**.”* (Barbosa, 1994, p.63)

O oclusiva nasal [J] é produzida, segundo este autor, com elevação do dorso da língua no alto do palato, produzindo aí uma oclusão.

Os resultados articulatórios provenientes de RM (2D), para um falante (Martins, 2007; Martins *et al.*, 2008) e dois falantes do PE (Oliveira, 2009), apontam para uma articulação mais anterior do que habitualmente descrito, apresentando uma área de contacto longa (desde o pré-palato até à região dental), o que está de acordo com o descrito, por exemplo, para o Catalão e outras línguas Românicas como atrás se referiu (ver secção 4.1.2.1).

No tocante a estudos acústicos, para esta consoante do PE, só muito recentemente (Monteiro, 2012) foi conduzida uma investigação que incluía falantes de dois dialetos do português (Bragança e Porto). Os valores de F1, F2 e F3 obtidos por Monteiro (2012) para a lateral palatal em posição intervocálica, foram respetivamente 300 Hz, 2050 Hz e 2900 Hz. Não se verificou a existência de diferenças significativas entre estes valores, para os dois dialetos considerados no estudo. Verificou-se, todavia, existir algum efeito de coarticulação uma vez que os valores das três primeiras formantes variaram em função do contexto vocálico considerado. De acordo com a autora, os resultados obtidos para o PE, relativamente a F1 e F3 estão em linha com os reportados na literatura para outras línguas. O valor de F2 foi próximo do obtido para o Espanhol e ligeiramente superior ao reportado por Silva (1996) para o PB. No entanto, no que diz respeito à duração da consoante este foi baixo (cerca de 80 ms), muito próximo ao obtido para a lateral alveolar e com a existência de diferenças entre os dois dialetos. Estes resultados contrariam a tendência reportada por Colantoni (2004) - segundo a qual a palatal é uma consoante mais longa que a lateral alveolar

- e são consideravelmente mais baixos que os obtidos, por exemplo, no estudo de Silva (1996) para o PB (198 *ms*).

4.2 Nasalidade

A nasalidade está presente em cerca de 99% das línguas do mundo, ou porque incluem vogais nasalizadas e/ou consoantes nasais (Chen *et al.*, 2007). O traço da nasalidade deve-se, essencialmente, ao abaixamento do véu palatino durante a produção dos sons, permitindo o escape de ar através das cavidades nasais. A descida do véu palatino condiciona um maior ou menor acoplamento entre o trato nasal e o trato oro-faríngeo.

Apesar da manobra articulatória requerida para a introdução de ressonâncias nasais no espectro acústico ser simples (abaixamento do véu palatino), as consequências resultantes deste acoplamento são difíceis de explicar, em virtude da complexidade geométrica do trato nasal como, por exemplo, a assimetria destas cavidades (Pruthi *et al.*, 2007). Por outro lado, é complicado separar os efeitos provocados pela abertura do porto velofaríngeo e as consequências resultantes de alterações na posição dos articuladores ao longo do trato oral (Feng e Castelli, 1996; Teixeira, 2000; Shosted *et al.*, 2012a).

4.2.1 O véu palatino e a passagem velofaríngea

O véu palatino ou palato mole é uma estrutura muscular e flexível que se prolonga a partir do palato duro. Na sua extremidade, apresenta uma estrutura móvel, a úvula que está envolvida na articulação de alguns sons como, por exemplo, as vibrantes uvulares (Mateus *et al.*, 2005).

O véu palatino é um dos articuladores mais importantes na implementação da nasalidade e tem, também, uma grande importância na delimitação da Passagem Velofaríngea ou Porto Velofaríngeo (PVF). A PVF é uma região deformável do trato nasal que estabelece a ligação entre a cavidade faríngea e o *cavum*. É delimitada posterior e lateralmente pelas paredes da faringe e anteriormente pelo véu palatino. Este quando elevado fecha a PVF. No tocante à produção dos sons da fala, a sua principal função é separar as cavidades nasais da cavidade oral.

Os movimentos do véu palatino e da PVF são controlados por cinco músculos, dois laterais superiores - músculo elevador do véu do palato (*levator veli palatini*) e o músculo tensor do véu do palato (*tensor veli palatini*) - dois laterais que se estendem inferiormente (palatoglosso e palatofaríngeo) e, ainda um outro músculo (central), o constritor superior da faringe (Serrurier, 2006; Serrurier e Badin, 2008).

O mecanismo de elevação do véu palatino é conseguido através da ação conjunta dos músculos elevador e tensor do véu palatino e do constritor da faringe. O primeiro contrai-se, elevando ou retraindo o véu palatino, o segundo cria tensão e alonga o véu palatino, puxando-o lateralmente. O constritor da faringe contribui para a subida do velo e tem um papel importante na deglutição. A descida do véu palatino fica a dever-se, em parte, ao relaxamento do tensor e do elevador, à ação dos músculos inferiores e à ação da gravidade. O palatoglosso (músculo comum

com a língua) é o principal responsável pelo abaixamento do véu palatino, sendo considerado o antagonista do músculo elevador do véu palatino. No entanto, o seu papel na produção de fala é alvo de controvérsia. O palatofaríngeo é também um dos músculos constituintes da parede da faringe, está implicado no movimento de descida do véu palatino, mas a sua função mais importante estará relacionada com o mecanismo de deglutição (Teixeira, 2000; Serrurier, 2006).

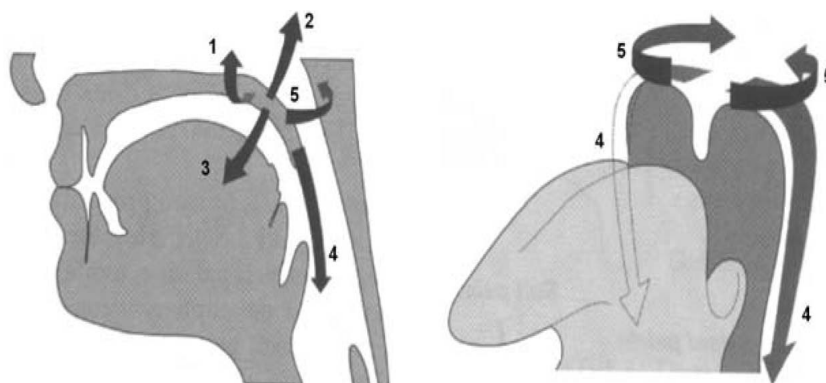


Figura 4.4: Músculos responsáveis pelos movimentos do véu palatino e da passagem velofaríngea. Tensor do véu palatino (1), Elevador do véu palatino (2), Palatoglosso (3), Palatofaríngeo (4) e Constritor da faringe (5). Fonte da imagem: Serrurier e Badin (2008).

4.2.2 Grau de abertura, altura do véu palatino

Como se referiu atrás, quando o véu palatino está baixo há acoplamento de cavidades com passagem do ar pelas cavidades nasais, permitindo a produção de sons nasais. Quando está elevado, impede a passagem do ar pelas cavidades nasais e são produzidos sons orais. Esta é, no entanto, uma visão demasiado simplificada da nasalidade, uma vez que o facto da PVF estar aberta, só por si, não implica que um som seja percebido como nasal ou que, estando fechada, não possam existir ressonâncias nasais. Para além disso, entre uma posição aberta e uma posição fechada, existem vários graus de acoplamento com graus de abertura intermédios e que variam na produção dos diferentes sons. O fenómeno da nasalidade será assim mais complexo do que a um primeiro olhar poderá parecer (Laver, 1980; Teixeira, 2000; Rossato *et al.*, 2006).

Vários estudos articulatórios demonstraram que é possível produzir sons não nasais com aberturas entre os 20–30 mm^2 . Por exemplo, algumas vogais orais, sobretudo vogais baixas como o [6], são produzidas com o véu palatino baixo, não sendo percebidas como nasais. Mesmo para as consoantes oclusivas, as consoantes em que o véu palatino terá que estar numa posição mais subida é possível ter aberturas de cerca de 10 mm^2 (Warren, 1964 referido por Teixeira, 2000).

Childers e Ding (1991), referidos por Teixeira (2000) afirmam que o grau de abertura do véu palatino varia em função do contexto fonético, encontrando-se aberto na produção dos sons nasais, numa posição intermédia para as vogais baixas, praticamente fechado para as vogais altas e completamente fechado na produção de oclusivas orais.

Moll (1960), num estudo cineradiográfico verifica que na produção das vogais baixas o

véu palatino apresenta um menor grau de fechamento, enquanto que na produção de vogais altas o grau de fechamento é maior. Observa também que as vogais adjacentes a consoantes nasais como o [n] apresentam fechamento incompleto e há tendência para um menor fechamento em vogais produzidas isoladamente do que em vogais inseridas em contexto de consoantes não nasais (e.g., [p] ou [k]).

Numa investigação realizada por Rossato *et al.* (2006), em que se estudaram algumas características de sons nasais e não nasais do PE e do Francês, no que diz respeito à altura do velo, os autores concluíram que, para ambas as línguas, a altura do velo segue a seguinte hierarquia: consoantes orais > vogais orais > consoantes nasais > vogais nasais.

Oliveira (2009), neste caso apenas para o PE, verificou que o véu palatino está, geralmente, numa posição mais elevada para os sons orais do que para os sons nasais, atingindo a sua maior amplitude na produção das vogais nasais. No que diz respeito às consoantes, esta autora refere que a posição mais elevada do véu palatino ocorreu para as fricativas surdas e a menor para as nasais. Alguns sons orais, por exemplo, as consoantes laterais, podem ser produzidas com o véu palatino parcialmente aberto, pelo menos em alguns contextos.

4.3 Vogais nasais

Na versão revista da *UCLA Phonological Segment Inventory Database* (UPSID), do total das 670 línguas consideradas no estudo (Maddieson, 2007), apenas 138 línguas (pouco mais de um quinto do total) apresentam no seu sistema vocálico uma ou mais vogais nasais. Ainda, de acordo com a mesma fonte:

1. nenhuma língua apresenta mais vogais nasais do que orais;
2. em 49.3% das línguas consideradas o número de vogais nasais iguala o número das vogais orais e, nas restantes, o número das vogais nasais é menor que o das vogais orais;
3. a maior diferença entre o número de vogais orais e nasais ocorre para o Francês que tem 11 vogais orais e apenas 3 vogais nasais.

A língua Portuguesa é, como já foi referido, extremamente rica em sons nasais, sendo uma das línguas Românicas em que as vogais nasais se apresentam como fonemas independentes (Sampson, 1999). As vogais nasais constituem uma das características distintivas do Português devido ao seu número e propriedades particulares (Oliveira e Teixeira, 2007b; Oliveira, 2009) representando, também, um dos pontos mais controversos da fonologia portuguesa (Moraes e Wetzels, 1992).

Para o Português, na sua variedade padrão, são consideradas, cinco vogais nasais: [i~], [e~], [6~], [o~] e [u~] tal como nas palavras **sim**, **penso**, **lã**, **dom** e **um**, respetivamente (Sampson, 1999; Mateus *et al.*, 2005). Tradicionalmente, as vogais nasais são classificadas como fechadas ou semi-fechadas, podendo no entanto existir variedades regionais com vogais nasais abertas ou semi-abertas. O [i~] e o [e~] são anteriores, o [6~] é central, o [o~] e [u~] são posteriores ou

também designadas por velares. A frequência de ocorrência destes fonemas do PE é, segundo Teixeira *et al.* (2001) de: 2.9% para o [6~], 1.2% para o [e~], 0.7% para o [i~], 0.9% para o [o~] e 0.6% para o [u~].

À semelhança do que acontece noutras línguas, o número de vogais nasais (cinco) do Português é muito menor do que o número de vogais orais (nove). Segundo Maddieson (2007), o conhecimento resultante de variados estudos acústicos tem demonstrado que é muito mais difícil para um ouvinte discriminar vogais nasais entre si do que as correspondentes vogais orais, o que pode, de certa forma, explicar porque é que, com muita frequência, o número de vogais nasais é menor do que o das vogais orais numa determinada língua. Este fator tenderá a ser preponderante em línguas em que o número de vogais é elevado, como é o caso do Português. O processo de evolução e a origem das vogais nasais poderá também explicar este facto, como se pode constatar no texto que transcrevemos:

“In the majority of languages with nasalized vowels the historical origin of these vowels can be traced to an earlier sequence of a nasal consonant followed by a vowel or a vowel followed by a nasal consonant “fusing” together. Such processes are quite likely to affect all vowels and thus produce nasalized counterparts to each oral vowel. However, if the number of vowels is large, maintaining the distinction between all the nasalized counterparts is difficult.” (Maddieson, 2007)

4.3.1 Breves notas sobre a evolução histórica das vogais nasais no Português e seu estatuto fonológico

No Português, apesar de as vogais nasais poderem nalgumas situações terem surgido por processos de assimilação progressiva, na maior parte das vezes, estas resultam de um processo de assimilação regressiva. A nasalização de uma vogal estabelece-se em duas etapas: primeiro, a vogal é nasalizada por efeito de uma consoante nasal que se lhe segue. De seguida, a consoante pode desaparecer, ficando uma vogal nasal fonémica. No caso da assimilação regressiva, a consoante nasal que desaparece encontra-se a seguir à vogal nasal. Os contextos mais comuns que resultam em vogais nasais, através de assimilação regressiva são:

1. /VNC/ - sequência de vogal que precede consoante nasal que por sua vez precede uma outra consoante;
2. /VN#/ - vogal precede uma consoante nasal final;
3. /VNV/ - vogal precede uma consoante nasal em posição intervocálica.

No contexto 1 e 2, a consoante nasal é tautosilábica i.e. pertence à mesma sílaba da vogal e forma coda silábica (e.g., têm-pus e pan). No terceiro contexto, a consoante nasal é heterosilábica com a vogal precedente como, por exemplo, na palavra lû-na (Sampson, 1999).

O estatuto fonológico das vogais nasais no português é, segundo alguns autores (Almeida, 1976; Moraes e Wetzels, 1992), um dos aspetos mais controversos da fonologia portuguesa,

dando espaço a diferentes correntes e interpretações. A principal questão reside no facto de se considerarem os cinco monotongos nasais como sendo representados por um ou dois fonemas, existindo ou não função distintiva. De seguida apresentam-se, de forma muito resumida, as diferentes interpretações:

1. Interpretação bifonémica - esta corrente será talvez a que reúne maior número de seguidores entre os estudiosos da língua portuguesa, quer para o PB (Cagliari, 1977b; Moraes e Wetzels, 1992) quer para o PE (Mateus, 1975; Almeida, 1976; Barbosa, 1961; Mateus e d'Andrade, 2000; Domingos, 2011). Esta interpretação não considera a existência de vogais nasais do ponto de vista fonológico. Estas são vistas como uma sequência de vogal e um segmento nasal /V+N/ (Domingos, 2011), um arquifonema nasal (Barbosa, 1961; Cagliari, 1977b) ou um autosegmento (Mateus e d'Andrade, 2000);
2. Interpretação monofonémica - esta corrente considera a vogal nasal na representação de base ou seja na matriz fonológica da língua. Assim, as vogais nasais serão fonemas distintos das vogais orais (Leite, 1979 referido por Moraes e Wetzels, 1992);
3. Outros autores consideram, fonologicamente, as vogais nasais como ditongos, sendo neste caso constituídas por um segmento oral e outro nasal (vogal). Esta visão é defendida por Parkinson (1983), tal como referido em Moraes e Wetzels (1992).

4.3.2 Aspetos dinâmicos e temporais da nasalidade vocálica no PE

Como já foi referido, o facto de durante a produção de um som nasal o véu palatino se encontrar numa posição baixa, permitindo o acoplamento entre os tratos oral e nasal parece não ser suficiente para explicar a complexidade acústica destes segmentos. Este acoplamento vai introduzir alguns zeros e pólos no espetro acústico, amortecendo os valores de frequência de algumas formantes, particularmente a amplitude da primeira formante (F1). No entanto, tem sido difícil encontrar pistas acústicas que caracterizem estes sons uma vez que existem muitos pólos/zeros que variam em função do grau de acoplamento (parâmetro difícil de avaliar) e as simulações efetuadas, considerando apenas o abaixamento do véu palatino, são incapazes de produzir sons nasais de elevada qualidade (Feng e Castelli, 1996; Teixeira, 2000).

Feng e Castelli (1996) consideram que ao gesto de abaixamento do véu palatino, na produção de uma vogal nasal, deve ser acrescentada uma perspetiva dinâmica:

“(...) the lowering of the velum makes the speaker pass from an oral configuration to that of an [N]-like nasal consonant, which can be seen as a target for nasal vowels (...) The nasalization of vowels is then considered as a dynamic trend from the oral configuration toward the target, although in the case of true nasal vowels it would not ever be reached.” (Feng & Castelli, 1996, p. 3695)

Reenen (1982), para além de referir a importância do grau de acoplamento nasal, e mais ainda, da proporção nasal/oral na perceção da nasalidade, refere que o nível de nasalidade, medido

através do N%, aumenta progressivamente durante a articulação de uma vogal nasal e considera este aspeto dinâmico crucial para a percepção da nasalidade vocálica.

A visão dinâmica da nasalidade já tinha sido também apontada para o português por Lacerda e Head (1966), na sequência de experiências com um pneumocromográfico que permite, em simultâneo, a obtenção das curvas de radiação nasal e oral.

“O decurso das vogais nasais e ditongos nasais é sempre oral-nasal com nasalidade médio-final ou final, excepto no caso de ocorrer uma consoante nasal anterior. Neste caso, o decurso é inteiramente nasal (...)” (Lacerda e Head, 1966, p.67)

Outra referência às questões dinâmicas e temporais na percepção da nasalidade vocálica, neste caso para o PB, foi efetuada por Clumeck (1976) que refere que mais importante que o gesto de amplitude vélica será o momento da sua abertura que melhor poderá contribuir para a percepção da nasalidade:

“(...)It might then be the case that the listener’s perception of the presence or absence of nasalization is more dependent upon the timing of palatal lowering rather than upon the actual extent of palatal lowering.” (Clumeck, 1976, p. 351)

Mais recentemente, utilizando síntese articulatória, Teixeira (2000) e Teixeira *et al.* (2000) verificaram que, para uma amplitude vélica adequada, a qualidade percetual das vogais nasais sintetizadas melhorava significativamente se se fizesse variar a posição do véu palatino ao longo do tempo, durante a produção de uma vogal nasal. Nestas experiências, foi considerada uma nasal produzida em três fases, em que a primeira fase tem um decurso oral, seguida por uma fase nasal propriamente dita e um murmúrio nasal¹¹ final.

4.3.2.1 As três fases da vogal nasal

O carácter dinâmico das vogais nasais permite identificar três fases, que têm sido atestadas através de vários estudos para as duas variantes do Português (PB e PE). No entanto, a existência de um início oral e do segmento nasal final não é consensual entre os vários autores e poderá depender do contexto em que a vogal é produzida e da própria posição que ocupa na sílaba. Relativamente à fase oral da vogal nasal, já Lacerda e Strevens (1956) lhe haviam feito referência após algumas experiências com um *speech-stretcher* e, posteriormente, com o pneumocromográfico:

“Nasal vowels in Portuguese have an initial segment whose degree of nasality varies vary greatly: In some cases the nasality may be so slight that in practice we can regard it as being absent; that is to say, we can regard the initial segment as being oral (...) but clearly some phonetic contexts are more likely to favour an initial oral segment than others are.” (Lacerda & Strevens, 1956, p.15-16 retirado de Lacerda e Head, 1966).

¹¹O murmúrio nasal é apenas um dos termos utilizados para fazer referência ao segmento consonântico nasal que pode surgir na fase final da vogal nasal. Outros termos como consoante nasal intrusiva, apêndice nasal, *nasal tail*, *excrecent nasal coda*, *nasal murmur*, entre outros são também utilizados. Neste trabalho, optou-se por utilizar o termo referido por cada um dos autores citados.

Silva (1995) mostrou que há uma primeira fase na produção da vogal nasal com posicionamento dos articuladores para pronúnciação da vogal oral correspondente com uma duração de 60–100 *ms*. A abertura da PVF dá origem a uma fase de transição com duração de 30–50 *ms* e, posteriormente, é possível identificar uma fase estável que corresponde ao acoplamento dos tratos oral e nasal.

Teixeira (2000), a partir de dados EMA verificou que mesmo em posição inicial, quando não existem quaisquer constrangimentos o véu palatino começa por subir e só, posteriormente começa a baixar.

Lovatto *et al.* (2007), com fibroscopia, num estudo para o PB confirma a existência de um murmúrio nasal e a existência de duas fases na produção da vogal nasal:

“Our results confirm that (i) an acoustic nasal tail N is clearly observed in 85% of Vn productions, (ii) when present, N is about half as long as the previous part of the vowel, and (iii) when Vn is in medial or final position, the maximum lowering of the velum is free to occur either before the nasal tail or during it, leading to much variability.”

Dois estudos acústicos relativamente recentes efetuados para o PE (e para outras variantes) apontam também para a existência de uma consoante nasal final, homorgânica com a oclusiva seguinte (Ashby *et al.*, 2011; Domingos, 2011).

4.3.2.2 Velocidade, duração dos gestos de abertura e fecho do véu palatino e coordenação entre gestos

O véu palatino é considerado um articulador relativamente lento, quando comparado com outros articuladores (e.g., ponta da língua). Os dados de cineradiografia reportados por Moll e Daniloff (1971) apontam para uma velocidade do véu palatino entre os 7 a 10 *cm/s*. Também, mais recentemente, num estudo de RM, Kim *et al.* (2010) fazem referência aos movimentos relativamente lentos do véu palatino, sendo possível captar os seus movimentos com uma resolução temporal de cerca de 126 *ms*. Para o caso da ponta da língua, uma resolução temporal adequada seria de 48 *ms* denotando diferenças de velocidades entre estes dois articuladores.

Vários autores referem que o ciclo completo de abertura/fecho do véu palatino dura entre 200 a 300 *ms* (Krakow, 1993; Stevens, 1998). A taxa de elocução, a posição silábica e o acento podem afetar, de forma significativa, a duração deste ciclo (Krakow, 1993). Uma taxa de elocução mais rápida, por exemplo, poderá encurtar estes tempos em cerca de 100 *ms*. Para Krakow (1993), a maior sobreposição de gestos poderá explicar esta diminuição do ciclo de abertura/fecho: *“as the rate increases, there is greater overlap among articulatory gestures for adjacent and near-adjacent segments”*, mas a existência de diferenças na velocidade dos dois movimentos pode ser também um fator a ter em consideração.

Para o Português, existem dados relativos à duração total do ciclo de abertura/fecho do véu palatino e para as várias fases do movimento (fase de abertura, fase estável, e fase de fecho) obtidos através de EMA (Oliveira e Teixeira, 2007b; Oliveira, 2009). Relativamente à duração

total, os autores obtiveram valores médios de duração total de cerca de 300 *ms* e verificaram que em taxa rápida o valor decresce para 200 *ms*. Estes resultados vão ao encontro dos referidos na literatura para outras línguas, como o Inglês ou Francês. De salientar que a fase de abertura do véu palatino foi, geralmente, mais longa que a de fecho e que cada uma destas fases é afetada pela taxa de elocução. Para as diferentes fases os valores médios encontrados foram 130-140 *ms* para a abertura, 30-40 *ms* para a fase estável e 110-120 *ms* para a fase de fecho. Esta diferença entre as durações de abertura e fecho parece não se verificar para o Francês, onde os movimentos têm uma duração semelhante (Amelot, 2004). Oliveira (2009) constatou que o fator posição silábica não afetava significativamente a duração total do ciclo de abertura e fecho. Relativamente ao efeito da vogal, os resultados não foram conclusivos tendo sido apontada a necessidade de obter informação para mais falantes (a informação reportada refere-se a um falante).

A velocidade dos movimentos do véu palatino foi também calculada por Oliveira (2009), utilizando o *stiffness*¹² como medida do movimento articulatorio. A autora verificou que o *stiffness* do fecho é sempre mais elevado que o de abertura e que este é influenciado pela taxa de elocução. O facto do *stiffness* de fecho ser mais elevado “*implica que o véu palatino seja mais lento a abrir do que a fechar*” (Oliveira, 2009, p.211).

Para além dos aspetos duracionais relacionados com os gestos do véu palatino, um aspeto que tem merecido algum destaque, particularmente à luz dos pressupostos da Fonologia Articulatoria, é a forma como os diferentes gestos (orais, nasais e glotais) se organizam temporalmente na produção das vogais nasais (Oliveira, 2009). São já alguns estudos que se focalizam nos aspetos de coordenação, particularmente, para o francês (Proctor *et al.*, 2013), PB (Oliveira e Marin, 2005) e PE (Oliveira e Teixeira, 2007b; Oliveira, 2009).

Para Oliveira (2009) o facto de, no português, o véu palatino atingir a sua máxima amplitude já na fase final da vogal nasal e o gesto de fecho do véu palatino ser posterior ao gesto oral é preponderante para a emergência da consoante nasal intrusiva ou murmúrio nasal. Também, como referido pela autora, percutualmente os ouvintes do Português preferem padrões de coordenação que favoreçam o seu aparecimento contrariamente aos ouvintes do Francês. Nesta língua o ponto de máxima abertura do véu palatino acontece sensivelmente a meio da vogal e quando o articulador oral atinge o *target*, o véu palatino já estará praticamente fechado.

4.3.3 A configuração oral das vogais nasais

A maior parte da literatura existente acerca da implementação das vogais nasais está quase sempre relacionada com a atividade do véu palatino (gesto nasal) e com os efeitos do acoplamento entre o trato oral e naso faríngeo. É frequentemente assumido que o contraste entre as vogais nasais e correspondentes vogais orais se deve ao abaixamento do véu palatino, sem que existam outros ajustes articulatorios significativos. No entanto, uma série de estudos articulatorios, efetuados para línguas em que as vogais orais contrastam fonologicamente com as vogais nasais apontam para a existência de ajustes articulatorios a nível da língua e dos lábios, mas esta poderá não ser

¹²Oliveira (2009) refere que o *stiffness* representa o quociente entre a velocidade máxima (*cm/s*) e o deslocamento máximo (*cm*) do articulador.

uma característica transversal às várias línguas (Carignan, 2011; Shosted *et al.*, 2012a).

O Francês será talvez a língua para a qual mais informação existe no que diz respeito à configuração oral das vogais nasais, proveniente de diversas técnicas instrumentais:

- Zerling (1984), tendo por base traçados de imagens radiográficas, verificou que, na língua francesa, para além do abaixamento do véu palatino, há consistentemente um movimento de recuo (retração da língua) e de labialização na produção das vogais nasais [A~] e [O~] do que nas correspondentes orais;
- Delvaux *et al.* (2002b) e Engwall *et al.* (2006) verificaram, utilizando imagens de RM, em falantes do Francês Belga, que a língua estava mais recuada e elevada em direção à zona velar na produção das vogais nasais. Esta configuração dos articuladores dará origem a um aumento da cavidade oral na zona frontal à constrição lingual. Estes ajustes na configuração oral podem ter como função preservar a qualidade vocálica compensando a presença das anti-formantes nasais, resultantes da abertura do PVF. No entanto, os autores sublinham que os diferentes falantes utilizam manobras articulatórias variáveis para efetuar o contraste entre vogais nasais e correspondentes vogais orais.
- Carignan (2011) corrobora os resultados prévios obtidos para o Francês confirmando que há ajustes orais a nível do dorso da língua e nos lábios.

Também, Shosted *et al.* (2012a), num estudo com dados articulatórios (EMA), acústicos e aerodinâmicos, para o Hindi¹³, reporta a existência de ajustes articulatórios orais na produção das vogais nasais em comparação às congêneres orais.

Para o Inglês, Carignan *et al.* (2011), num estudo articulatório (EMA) e aerodinâmico, verificaram que o [i~] é produzido com o dorso da língua numa posição mais elevada que a vogal oral [i] podendo esta diferença ser encarada como uma estratégia para compensar o desvio espectral nas baixas frequências originado pela abertura da PVF. Para o par [a/a~] não se verificou nenhum mecanismo compensatório a nível oral.

Para o PB, Master *et al.* (1991), compararam os traçados provenientes de xeroradiografia obtidos para as vogais orais e nasais para um falante, e verificaram que as diferenças articulatórias entre o [i] e o [i~] eram mínimas mas que para o par [a] e [a~] existiam ajustes orais significativos.

Matta Machado (1993), com dados de cineradiografia, num falante carioca (Rio Janeiro) conclui que, para além do abaixamento do véu palatino, existe na produção das vogais nasais: redução da cavidade bucal; redução da cavidade faríngea para as vogais [i~], [e~] e [o~] e aumento desta para o [6~] e [u~]; ligeira contração da parede da faringe e uma duração maior que as correspondentes orais.

Mais recentemente, Gregio (2006), para um falante, utilizando RM reportou diferenças a nível da posição da língua entre as vogais nasais posteriores e as suas equivalentes orais, obedecendo a magnitude destes ajustes à progressão: [6~] > [o~] > [u~]. No caso do [6~] o dorso da língua

¹³O Hindi tem dez vogais orais e dez vogais nasais e estas contrastam fonologicamente (Shosted *et al.*, 2012a).

está mais anteriorizado do que na oral [a] e, no caso das nasais [u~] e [o~], o dorso da língua está mais retraído do que nas correspondentes orais com efeitos a nível da cavidade faríngea.

Medeiros e Demolin (2006), também com RM em tempo real, verificaram que existem pequenos ajustes articulatórios orais para o [6~], em que a parte posterior da língua parece acompanhar o movimento do véu palatino. A parte anterior da língua está mais elevada em [i~] do que no [i] e apresenta-se mais plana no [u~]. Segundo os autores, estas diferenças de posicionamento da língua serão mecanismos de compensação para criar conformações de ressonância no tubo que permitam a produção da qualidade vocálica adequada.

No tocante ao PE, a maior parte dos estudos sobre nasalidade vocálica tem sido focalizado no estudo da dinâmica do véu palatino (Teixeira e Vaz, 2001; Rossato *et al.*, 2006; Oliveira e Teixeira, 2007b; Oliveira *et al.*, 2009) e pouco na configuração oral destes sons. A única informação que existe foi obtida através de RM 2D, através de configurações estáticas, por Martins (2007); Martins *et al.* (2007) para um falante e, posteriormente, por Oliveira (2009) para três falantes do PE.

Nos estudos de Martins (2007) e Martins *et al.* (2007) verificou-se, através da sobreposição de contornos para vogais orais e nasais que, para as vogais [i~], [e~] e [u~], não existiam ajustes articulatórios orais significativos, para além do abaixamento do véu palatino. No caso da vogal nasal [6~], esta foi produzida com uma posição do dorso da língua mais elevado e mais avançado do que as orais [6] ou [a]. Constatou-se que não existiam ajustes faríngeos dignos de referência, particularmente ao nível da sua parede posterior, o que vai ao encontro do descrito por Delattre (1968), que refere que contrariamente ao Francês, a nasalidade vocálica no português não envolve guturalização, passando apenas por um abaixamento do véu palatino. No caso particular do [6~] as dimensões da cavidade faríngea, pelo menos em 2D, são maiores na vogal nasal do que na congénere oral o que está de acordo com as descrições efetuadas, para o PB, por Matta Machado (1993) e Gregio (2006).

Os resultados obtidos por Oliveira (2009), para um maior número de falantes, corroboram as observações previamente efetuadas para o PE, concluindo que, no geral, existem pequenos ajustes articulatórios orais (maiores para as vogais [6~] e [o~]), cuja magnitude é, todavia, muito menor do que o reportado para a língua francesa. No entanto, o facto das articulações serem sustidas, nos estudos efetuados para o PE, poderá não captar os aspetos dinâmicos da nasalidade vocálica nesta língua.

Capítulo 5

RM Aplicada ao Estudo das Consoantes Laterais do Português

One limitation of midsagittal models is their inherent inability to characterize lateral consonants, as laterals are characterized by the presence of a complete closure in the midsagittal plane while lateral channels are maintained open. The detailed 3D knowledge of the vocal tract shape is important to deal in a more realistic way with the aerodynamic models needed for speech production studies.

BADIN *et al.* (2002)

Divide as dificuldades que tenhas de examinar em tantas partes quantas for possível, para uma melhor solução. Não existem métodos fáceis para resolver problemas difíceis.

RENÉ DECARTES

No presente capítulo congregar-se-á toda a informação relativa à recolha e processamento dos dados, apresentação e discussão dos resultados conseguidos para as consoantes laterais. Assim, começaremos por enquadrar o tema fazendo uma breve referência a estudos semelhantes, vocacionados para o estudo destes sons, realizados para outras línguas. De seguida, apresentam-se as questões que pretendemos vir a abordar e os objetivos que pretendemos alcançar com esta parte do nosso trabalho que envolve a aquisição e processamento de informação 2D e 3D obtida a partir de produções sustentadas. Procederemos depois à descrição do método onde se incluirá informação acerca dos participantes no estudo, do *corpus* selecionado, dos procedimentos experimentais relacionados com a aquisição de imagem, com o processamento dos dados e com algumas tarefas de avaliação (preliminares) efetuadas. Na parte seguinte do capítulo reportar-se-ão os resultados alcançados com o presente trabalho e que permitirão efetuar a caracterização articulatória das consoantes laterais. Finalmente, as implicações destes resultados no conhecimento fonético serão alvo da discussão.

5.1 Estudos RM para estudo e modelação das consoantes laterais

Apesar de serem muitos os estudos que recorrem à utilização de RM na área da produção de fala (veja-se secção 2.2 do capítulo 2) não são muitos os estudos, mesmo para outras línguas, que se focalizam na caracterização das consoantes laterais¹.

Da pesquisa efetuada verificamos que a maior parte dos estudos com RM centrados no estudo da lateral alveolar - /l/ foram efetuados para o Inglês, particularmente, para o Inglês Americano (Bangayan *et al.*, 1996; Story *et al.*, 1996; Narayanan *et al.*, 1997a; Ong e Stone, 1998; Zhang *et al.*, 2003; Zhang e Espy-Wilson, 2004; Zhou, 2009). Destes, a maior parte teve como objetivo principal efetuar modelação acústico-articulatória (Bangayan *et al.*, 1996; Zhang *et al.*, 2003; Zhang e Espy-Wilson, 2004; Zhou, 2009) não só do /l/ mas também de outras consoantes líquidas (e.g., o [4]).

Ong e Stone (1998) apresentam a configuração 3D do trato vocal para a lateral /l/ e para os sons róticos utilizando, para além de imagens de RM, dados de US, EPG e dados acústicos.

Story *et al.* (1996), para um falante de inglês, incluem no *corpus* uma produção do /l/ (em *lump*) no conjunto de dezoito configurações do trato vocal (entre vogais e consoantes) adquiridas com RM (26 cortes de 5 mm em 4 m e 15 s). Para estimar o espaço ocupado pelos dentes, na cavidade oral, os autores utilizaram imagens destas estruturas obtidas com TC por feixe de electrões, do mesmo falante, num estudo prévio. Neste caso, o objetivo do estudo foi bastante abrangente não sendo direcionado para o /l/. No que diz respeito a esta consoante são apresentadas as áreas ao longo do trato vocal, o valor do comprimento do trato vocal e uma visualização 3D para ilustrar o estabelecimento das passagens laterais.

Dos estudos acima mencionados merecem-nos destaque, por serem os que incluem uma descrição mais detalhada da lateral alveolar, os trabalhos de Narayanan *et al.* (1997a) e, mais recentemente, o trabalho de Zhou (2009). Este último é importante na medida em que inclui vasto trabalho de modelação acústico-articulatória do /l/.

O trabalho com laterais reportado em Narayanan *et al.* (1997a) representa parte de uma vasta linha de investigação em que a RM é utilizada para efetuar modelação tridimensional de várias classes de sons do inglês. No conjunto foi efetuada caracterização articulatória de vogais em Narayanan *et al.* (1997b), de consoantes fricativas (Narayanan *et al.*, 1994, 1995) e dos sons róticos (Alwan *et al.*, 1997). Os falantes, o protocolo de aquisição e as técnicas de processamento utilizadas foram transversais aos vários estudos.

No estudo de Narayanan *et al.* (1997a) os autores recorreram a uma sequência *EG Spoiled GRASS*, com espessura de corte de 3 mm. Foram obtidos cortes coronais (40 a 45 cortes), axiais (40 a 45 cortes) e sagitais (28 a 35 cortes) para obtenção de informação 3D. Para conseguir cada umas das configurações do trato vocal houve necessidade do falante repetir o som, seis a nove

¹Existem, no entanto, variadíssimos estudos que utilizam outras técnicas articulatórias como Cine-radiografia, EMA, EPG, ou US para obtenção de dados articulatórios para /l/ e também, em menor número, para o /L/. No capítulo precedente, já foi efetuada referência a alguns desses estudos pelo que nos escusaremos, agora, de os repetir.

vezes, por períodos de 13 a 16 s. A aquisição de imagens em cada um dos planos (e.g., axial) para o mesmo sujeito foi efetuada na mesma sessão, sendo o conjunto dos três planos adquirido ao longo de três sessões distintas e em dias separados. A informação fornecida, relativa à forma como foi obtida a informação das arcadas dos falantes, deixa entender que foi um processo complexo com obtenção de imagens RM em moldes feitos de *Alginate Plus* e posterior criação de *templates* que guiaram o processo de segmentação. No entanto, não são apresentadas imagens onde sejam observados os dentes.

O *corpus* foi obtido com quatro falantes (dois do sexo feminino e dois do sexo masculino), todos com treino fonético. As variantes (*dark* e *light*) da lateral /l/ foram produzidas no contexto de vogais neutras. Neste estudo, cujo principal objetivo era descrever as características 3D das variantes *light* e *dark* do /l/ são apresentadas medidas do comprimento do trato vocal, funções de área do trato vocal e visualizações 3D da língua. A informação relativa ao ponto e extensão da articulação é complementada com dados de EPG. Os autores realçam a importância da informação 3D no estudo destes sons e identificam pontos comuns na articulação das duas variantes do /l/ como a compressão lateral da língua, forma convexa do seu dorso e presença de contacto alveolar.

O estudo de Zhou (2009) incluiu apenas um falante², tendo sido analisada a configuração do trato vocal para um *dark-l* e um *light-l* obtidas a partir de produções sustentadas. As imagens de RM do falante foram selecionadas de uma base de dados recolhida, anteriormente, na Universidade de Cincinnati por Mark Tiede (Tiede, 2004 referido por Zhou, 2009). Inclui imagens ponderadas em T2 (SS-FSE) no plano sagital (3 cortes) e imagens coronais e axiais (24 a 33 imagens), com 5 mm espessura, adquiridas com sequência EG-Spoiled (Fast Multi Planar Spoiled Gradient Echo – FMPSRGR). O tempo de aquisição foi de 25 s. No que diz respeito às arcadas dentárias é referido que estas foram obtidas a partir de imagens de TC não ficando claro se foram obtidas a partir de moldes dentários do falante ou diretamente numa aquisição com o falante. Não são mostradas imagens com as arcadas co-registadas. Foi efetuada modelação por elementos finitos (FEM) do trato vocal e foi utilizado o *software* VTAR (*Vocal Tract Acoustic Response*) para efetuar modelação acústico-articulatória. O trabalho no que diz respeito à lateral centra-se nos efeitos das passagens laterais e da cavidade supra-lingual no espetro acústico.

Para o inglês britânico, Gick *et al.* (2002), obtiveram imagens RM 2D referentes a vogais e consoantes líquidas (/l/ e [4]) para estabelecer comparação entre os gestos articulatórios destas consoantes e os de algumas vogais. Foram obtidas as distâncias médio sagitais ao longo do trato vocal. O estudo não contempla informação 3D.

Finalmente, caberá referir um estudo para o Tamil, uma língua Dravidiana falada no sul da Índia, que conjuga dados articulatórios de RM, EMA e EPG e análise acústica (Narayanan *et al.*, 1999). Esta língua apresenta cinco consoantes líquidas contrastivas, algo pouco usual dentro das línguas conhecidas, sendo que duas são consoantes laterais. No estudo é apresentada uma descrição pormenorizada das consoantes líquidas, são enfatizadas as diferenças entre as laterais e as outras consoantes líquidas e são identificadas as características específicas das laterais.

²Esta investigação para além da modelação acustico-articulatória para a lateral alveolar (*dark-l* e *light-l*) incluiu trabalho semelhante para o *retroflex* e *bunched* /r/.

Para o caso do /L/, que seja do nosso conhecimento, não existe nenhum estudo com RM que se focalize numa caracterização detalhada desta consoante.

5.2 Caracterização do problema e questões a investigar

Como tivemos oportunidade de referir, os estudos articulatórios (e mesmo acústicos), vocacionados para o estudo das consoantes laterais, para o PE, são escassos. Para além da escassez de dados, são também controversas as posições de diferentes autores quanto a alguns aspetos relativos à produção das laterais, particularmente, da lateral alveolar (Oliveira, 2009).

O estudo das consoantes laterais e a sua modelação acústico-articulatória são dificultadas pela sua complexa geometria. A informação bidimensional, apesar de importante, parece ser insuficiente para caracterizar sons em que alguns dos acontecimentos articulatórios ocorrem para lá da linha média, sendo fundamental a obtenção de informação 3D (Badin *et al.*, 2002; Narayanan *et al.*, 1997a; Zhou, 2009).

Como ficou já expresso no capítulo introdutório, um dos objetivos gerais deste trabalho visa a descrição articulatória destes sons, tendo em vista contribuir para um melhor conhecimento da sua produção e permitir o desenvolvimento de um modelo articulatório que acomode a síntese dos sons laterais. Este modelo deverá contar com a informação e com os dados articulatórios que nos propomos recolher. Para que tal seja possível torna-se necessário clarificar e enquadrar algumas questões mais teóricas, ainda não totalmente esclarecidas, ou polémicas relativamente à produção destes sons.

Assim, as questões que pretendemos retomar e investigar, utilizando a RM como forma de obter informação, são as seguintes:

1. Como se caracterizam os gestos articulatórios associados à lateral alveolar, no que diz respeito à ponta da língua (localização e extensão da constrição, passagens laterais)?
2. Qual a influência da posição silábica na articulação do /l/?
3. A configuração articulatória da lateral alveolar é compatível com a de um /l/ velarizado?
4. De que forma o /l/ é influenciado pelo contexto vocálico?
5. Qual a configuração articulatória do /L/, i.e. quais os gestos associados à lateral palatal /L/?

5.2.1 Objetivos do estudo

Os nossos objetivos concretos a atingir nesta parte do nosso trabalho são:

- Elaboração de um *corpus* específico para os sons laterais que permita averiguar o efeito da posição silábica e do contexto vocálico, para que se possa dar um contributo para a clarificação das questões atrás enunciadas;

- Optimização de protocolos adequados para a aquisição de uma base dados de RM (2D e 3D) com informação proveniente de vários falantes;
- Explorar diferentes abordagens de processamento e segmentação para extração da informação;
- Obtenção de dados articulatórios de natureza qualitativa (comparação contornos) e quantitativa como, por exemplo, áreas ao longo do trato, dimensão das passagens laterais, etc. que permitam a caracterização articulatória das consoantes laterais.

5.3 Método

Na presente secção, efetuar-se-á uma descrição, tão detalhada quanto possível, de todas as metodologias seguidas, para a obtenção de dados articulatórios relativos às laterais. Começaremos com a descrição do *Corpus*, dos falantes, das sessões experimentais, da recolha de dados (RM e acústicos) e das técnicas de processamento de imagem utilizadas neste estudo estático (2D e 3D).

5.3.1 *Corpus*

Ainda que o propósito deste trabalho seja a obtenção de dados articulatórios para as consoantes laterais o *corpus* incluiu material linguístico relativo a outros sons para os quais a informação disponível para o PE também é escassa como, por exemplo, os róticos ou vogais centrais³.

Para a lateral alveolar /l/ foram consideradas três das posições silábicas em que esta pode ocorrer no PE: 1) ataque (início de palavra), 2) coda (posição final) e 3) ataque intervocálico. Outras posições em que o /l/ pode ocorrer, como, por exemplo, o ataque complexo ou o ataque ramificado não foram incluídas no *corpus*, pois, apesar de ser desejável explorar todos estes contextos, um compromisso teve que ser estabelecido para que o tempo total de aquisição, por falante, fosse aceitável. Para avaliar a influência do contexto vocálico, a consoante lateral alveolar foi adquirida no contexto das vogais cardinais do PE - [i, a, u], nas três posições consideradas. As palavras de referência selecionadas para instruir o falante foram retiradas (com algumas modificações de pormenor) dos *corpora* elaborados para recolha de dados EMA para o PE (Oliveira e Teixeira, 2007a; Teixeira e Vaz, 2001). A par com o presente trabalho, foi também adquirido um *corpus* acústico para estudo da lateral-alveolar /l/, mais extenso do que o de RM, utilizando as mesmas palavras e contextos (Marques, 2010). O objetivo é poder comparar, quando se justifique, dados provenientes das diferentes técnicas.

Para a lateral dorso-palatal /L/, apenas foi considerada a posição intervocálica por ser a que mais frequentemente ocorre no PE⁴. As palavras de referência usadas foram: **palha**, **pilha**

³Para os sons róticos ou consoantes vibrantes foram adquiridas a vibrante alveolar [r] e a vibrante múltipla [R] que se opõem em posição intervocálica. Foram utilizadas as seguintes palavras de referência: **caro** [ka4u] (posição intervocálica), **mar** [ma4] (coda final absoluto) e **carta** [ka4t6] (final de sílaba) para a vibrante simples e, para a vibrante múltipla as palavras **carro** e **rato**. As vogais centrais adquiridas, para além do [a] foram o [ɪ] como em **devi** [d1vi] e [ɛ] como em **cada** [k6d6].

⁴De acordo com Mateus e d'Andrade (2000).

e **pulha**. A tabela 5.1 contém todas as palavras de referência selecionadas para elicitare os sons. Para cada palavra é apresentada a transcrição fonética, em alfabeto X-SAMPA.

Tabela 5.1: *Corpus de laterais adquirido com RM (2D e 3D)*

Fone	Ataque	Coda	Intervocálica
/l/	laca [lak6]	sal [sal]	sala [sal6]
	litro [lit4u]	til [til]	bilis [biliS]
	lupa [lup6]	sul [sul]	pulo [pulu]
/L/			palha [paL6]
			pilha [piL6]
			pulha [puL6]

O *corpus* estático contemplou ainda as vogais cardinais [a, i, u], elicitadas a partir das palavras de referência: **pato**, **bica**, e **buda**. As vogais cardinais foram incluídas, uma vez que os sons laterais foram adquiridos em contexto vocálico, sendo necessário (para comparação) ter informação relativa às vogais, para os mesmos falantes.

5.3.2 Falantes/Informantes

Com o intuito de selecionar candidatos para participar no estudo, foi efetuada divulgação junto de alunos do curso de Radiologia, Terapia da Fala, colegas e outros investigadores. Os que se mostraram interessados em participar foram sujeitos a uma avaliação mais detalhada, com o intuito de verificar o cumprimento dos critérios de inclusão/exclusão que passamos a enumerar:

Critérios de inclusão

- Falante nativo do Português Europeu de qualquer naturalidade (preferencialmente de Portugal Continental)
- Desenvolvimento cognitivo normal
- Classe Oclusal do Tipo I

Critérios de exclusão

- Ser menor de 18 anos
- Alterações/patologia a nível de linguagem, fala e audição
- Presença de lesão neurológica evidente
- Gravidez
- Presença de contra-indicações à realização de RM (e.g., pace-maker, clips de aneurisma craniano, implantes cocleares)

- Presença de próteses ou tratamentos dentários que, mesmo sendo seguros num ambiente de RM, pudessem dar origem a artefactos, com degradação da qualidade de imagem

Numa fase inicial, efetuou-se um primeiro rastreio dos candidatos, para excluir a presença de objetos metálicos a nível da cavidade oral ou na zona do pescoço, que mesmo não sendo impeditivos da realização da RM poderiam afetar a qualidade das imagens. A presença de coroas, implantes, aparelhos de correção e fios de fixação inviabilizou, logo à partida, a participação de alguns candidatos. Todos os potenciais candidatos efetuaram seguidamente uma sessão com um terapeuta da fala experiente para detetar possíveis alterações a nível da linguagem e da fala e para determinação da classe oclusal.

Numa fase posterior, todos os sujeitos participantes no estudo foram informados, através da leitura de uma ficha de informação, dos objetivos do estudo, das condições de realização do exame de RM, dos riscos e das questões de confidencialidade (ver anexo A). Foi também preenchida uma ficha de recolha de dados (cf. anexo B), com o intuito de permitir uma melhor caracterização dos participantes quanto a idade, constituição física (peso e altura), nacionalidade, naturalidade, nível académico e nível de conhecimentos de fonética /linguística.

Depois deste processo de seleção, participaram neste estudo sete sujeitos adultos (três do sexo feminino e 4 do sexo masculino) com idades compreendidas entre os 20 e os 39 anos. Todos participaram no estudo de forma voluntária e gratuita. Nenhum dos informantes apresenta treino vocal específico e o nível de conhecimentos de fonética é variável. Um dos falantes (JH) é bilingue (falante do PE e de Espanhol da América Sul).

Apresenta-se de seguida uma caracterização mais detalhada de cada um dos falantes identificados através das siglas correspondentes aos nomes, respeitando o acordo de confidencialidade efetuado entre o autor e os participantes.

1. **CO** - informante do sexo feminino, com 31 anos, natural de Viseu, residente na cidade de Aveiro há 7 anos. Habilitações académicas a nível do Doutoramento. Tem conhecimentos de fonética e linguística. Mede 160 cm e pesa cerca de 50 Kg.
2. **ER** - informante do sexo feminino, com 22 anos, natural e residente em Pardilhó (distrito de Aveiro). Finalista da Licenciatura em Radiologia da ESSUA. Não tem conhecimentos de fonética ou linguística. Mede 150 cm e pesa cerca de 50 Kg.
3. **MC** - informante do sexo feminino, com 21 anos, natural e residente em Oliveira do Bairro (distrito de Aveiro). Finalista da Licenciatura em Radiologia da ESSUA. Não tem conhecimentos de fonética ou linguística. Mede 160 cm e pesa cerca de 63 Kg.
4. **LCR** - informante do sexo masculino, com 21 anos, natural e residente no Cartaxo (distrito de Santarém). Aluno do 3º Ano da Licenciatura em Radiologia da ESSUA. Não tem conhecimentos de fonética ou linguística. Mede 175 cm e pesa cerca de 75 Kg.
5. **JH** - informante do sexo masculino, com 39 anos, natural da Venezuela e residente na zona de Coimbra desde os 2 anos de idade. Bilingue (PE e Espanhol). Licenciado em Audiologia e

Mestre em Ciências das Fala e Audição. Tem conhecimentos de fonética e linguística. Mede 183 cm e pesa cerca de 85 Kg.

6. **JPM** - informante do sexo masculino, com 33 anos, natural de Tomar (distrito de Santarém) e residente em Aveiro. Licenciado em Radiologia. Não tem conhecimentos de fonética ou linguística. Mede 190 cm e pesa cerca de 98 Kg.
7. **AS** - informante do sexo masculino, com 21 anos, natural e residente em Ansião (distrito de Coimbra). Aluno do 3º Ano da Licenciatura em Radiologia da ESSUA. Não tem conhecimentos de fonética ou linguística. Mede 181 cm e pesa cerca de 75 Kg.

5.3.3 Recolha de dados

5.3.3.1 Local e Período

A aquisição das imagens de RM foi contratualizada através da Rede de Imagiologia Funcional (via Instituto Biomédico de Investigação em Luz e Imagem - IBILI), em Coimbra. A aquisição das imagens de RM foi efetuada em 3 sessões distintas (de três horas cada), durante os meses de Abril e Maio de 2010, com o suporte técnico considerado necessário e fornecido pela instituição. Os dados foram, posteriormente, processados no Instituto de Electrónica e Telemática da Universidade de Aveiro (IEETA).

5.3.3.2 Equipamento e material

As imagens de RM foram, na sua totalidade, adquiridas num equipamento com intensidade de campo de 3T (Magnetom TIM TRIO, Siemens, Erlangen, Alemanha) equipado com gradientes de elevada performance (Máxima amplitude - 45 mT/m , *rise time* - $200\text{ }\mu\text{s}$ e *slew rate* - 200 T/m/s , *FOV* - 50 cm). Foram utilizadas, em simultâneo, as antenas *phased-array* de crânio (12 canais) e de pescoço (4 canais) com recurso a imagem paralela (*GRAPPA*).

5.3.3.3 Sessões preparatórias

O trabalho de otimização das sequências e procedimentos seguidos no presente estudo decorreram nos meses anteriores às recolhas de dados. Como já foi referido, o trabalho efetuado pela autora, no âmbito do Mestrado em Ciências da Fala, envolveu também a recolha de imagens de RM no contexto da produção de fala. Desta forma, o trabalho de escolha e otimização das sequências para estabelecer o protocolo final de aquisição 3D para o *corpus* de laterais foi mais facilitado. No entanto, houve ainda que proceder a pequenos ajustes uma vez que as aquisições de imagem (no Mestrado e para o presente trabalho) não decorreram na mesma instituição e os equipamentos apresentam intensidade de campo magnético diferente (1.5 T e 3.0 T), o que implicou alterações nos parâmetros das sequências. Foram testadas algumas sequências (e.g., 3D SPACE - *Sampling Perfection with Application optimized Contrasts using different flip angle Evolutions* e 3D MP-RAGE) mas constatou-se que a sequência EG - *Spoiled* 3D VIBE (já utilizada no estudo anterior)

foi a que apresentou o melhor compromisso entre o tempo de aquisição e a qualidade global das imagens, tendo em atenção os nossos objetivos. Para além disso, com esta sequência os constrangimentos com o SAR são menores uma vez que é uma sequência EG e utiliza FA relativamente baixo.

Neste estudo, como o objetivo se centra no estudo das consoantes laterais e como se pretendia melhor visualização da cavidade oral (para identificação das passagens laterais) optou-se por fazer a aquisição do volume de dados no plano sagital. No estudo anterior, a aquisição do volume foi efetuada no plano axial o que, na situação atual, não seria a melhor escolha. A aquisição do volume no plano coronal, por outro lado, obrigaria à utilização de FOV e espessura da *slab* maior, o que teria impacto na resolução espacial ou no tempo de aquisição. No que diz respeito ao tempo de aquisição, teria sido viável encurtá-lo um pouco, todavia, optou-se por não o fazer uma vez que a degradação da qualidade das imagens poderia, posteriormente, dificultar as tarefas de processamento e, em consequência, aumentar a dificuldade na discriminação das passagens laterais.

Foi durante este período preparatório que se fizeram algumas experiências no sentido de encontrar uma solução que permitisse efetuar, *a posteriori*, o co-registo das arcadas dentárias dos falantes. São conhecidas as limitações da RM na visualização de estruturas ósseas e calcificadas (e.g., dentes) devido à falta de prótons de Hidrogénio⁵. É reconhecida a necessidade de identificar o espaço ocupado pelos dentes na cavidade oral para que se consiga discriminar o fluxo de ar lateral que se forma dos lados da língua na produção das consoantes laterais. Desta forma, teve que se encontrar a melhor solução para obter essa informação.

Várias têm sido as abordagens seguidas pelos investigadores para ultrapassar este problema. Em alguns estudos, os falantes efetuam uma aquisição volumétrica de alta resolução das arcadas recorrendo a TC e estas são, posteriormente, co-registadas com as imagens de RM (veja-se Serrurier e Badin, 2008 num estudo de modelação para vários articuladores). Outra estratégia, que evita a exposição do falante à radiação, é efetuar um molde dos dentes do falante e adquirir imagens de TC do molde (cf. Zhou, 2009). Alternativamente, pode-se colocar o molde das arcadas dentro de água e obter imagens RM (Narayanan *et al.*, 1995). Takemoto *et al.* (2004) propõem uma aquisição de RM dirigida às arcadas do falante, com este em posição de pronação, com a cavidade oral preenchida com sumo de mirtilos. Estes frutos apresentam propriedades paramagnéticas induzindo um efeito semelhante ao dos produtos de contraste (e.g., gadolínio) utilizados em RM, permitindo a discriminação dos dentes.

No nosso estudo em 2007 testou-se um método semelhante ao descrito por Takemoto *et al.* (2004), colocando água na boca do falante com aquisição de sequência de alta resolução

⁵Caberá notar que quando se refere que, nas imagens RM, a visualização dos dentes, mandíbula e outras estruturas calcificadas é má, inferior à proporcionada por RC ou TC, geralmente consideraram-se as técnicas e sequências convencionais que são utilizadas, na prática diária, na maior parte dos setores de RM e que são as que utilizamos também neste estudo. Note-se, no entanto, que atualmente as potencialidades da RM, em aplicações do foro dentário, têm vindo a aumentar começando a revelar-se muito interessante a sua utilização neste contexto. A utilização de algumas sequências/técnicas como *ultrashort TE (UTE)*, *sweep imaging with Fourier transformation - SWIFT*, a utilização de antenas dedicadas (e.g., intra-orais) em campos de intensidade elevada podem proporcionar visualização destas estruturas com muito boa qualidade. Uma vez que o assunto sai um pouco do contexto deste trabalho, remetemos a leitor para alguns artigos, recentemente publicados, que ilustram o potencial da RM nesta área: Tymofiyeva *et al.* (2008), Idiyatullin *et al.* (2011) ou Idiyatullin *et al.* (2013).

ponderada em T2 (TSE 3D). Nesta situação, a água apresenta hipersinal permitindo a visualização do contorno dos dentes (a preto).

No contexto deste trabalho, a aquisição de imagens de TC nos falantes nunca se afigurou como alternativa uma vez que envolve o recurso a radiações ionizantes o que num estudo desta natureza, efetuado em voluntários jovens sem qualquer patologia, não seria justificável nem uma boa prática.

Optar por fazer moldes dentários para cada um dos sete falantes, efetuar aquisição de TC ou RM aos moldes e fazer co-registo com os volumes de RM obtidos em fonação, implicaria custos avultados com materiais dentários e consultas de estomatologia, dado o número de falantes, que não poderíamos suportar.

O recurso ao método que se testou no nosso estudo prévio seria uma possibilidade, todavia, segundo reportado pelo falante, a experiência não foi muito confortável pois implicou estar cerca de três minutos em posição de pronação com a boca preenchida com água. Para além disso, na instituição onde decorreram as aquisições de imagem para o presente estudo, seria necessário o recurso a uma antena de corpo uma vez que a antena crânio, dada a sua configuração, impede a colocação do falante em posição de pronação. Ter que manter água, sumo de mirtilos ou outro líquido dentro da boca, durante alguns minutos, em posição de supinação, não nos pareceu uma hipótese a considerar.

Surgiu-nos a ideia de preencher goteiras dentárias (de material plástico) com compota de mirtilos e colocá-las dentro da boca do falante. Foram efetuadas algumas aquisições de imagem que nos possibilitaram identificar o contorno dos dentes (a preto) contrastando com o sinal mais elevado da compota à sua volta. No entanto, a dimensão das goteiras e a salivagem, resultante da presença da compota dentro da boca, revelou-se desconfortável para o falante.

Ao analisar as imagens adquiridas com o 3D VIBE durante as nossas experiências e considerando os resultados obtidos num estudo que visava comparar as medidas efetuadas sobre reconstruções da mandíbula (*in vitro*) a partir de aquisições RM-3D e TC⁶ (veja-se Goto *et al.*, 2007) pareceu-nos viável extrair a informação das arcadas utilizando a sequência 3D VIBE, sem ser necessário recorrer a qualquer meio de contraste.

Dada a orientação destas estruturas optou-se pela aquisição da sequência no plano axial de forma a incluir toda a mandíbula e maxilar superior. Instruiu-se o participante para colocar a língua entre os incisivos superiores e inferiores, permitindo a sua separação, e para manter saliva dentro da boca durante a aquisição. A qualidade das imagens obtidas indicou a possibilidade de se poder extrair a informação de que necessitávamos relativa às arcadas.

As imagens obtidas, nestes testes, serviram-nos para ensaiar o método de segmentação (descrito adiante na secção 5.3.4.1) e avaliar se, de facto, seria possível obter a informação desejada

⁶No estudo referido foram testadas três sequências (e.g., True FISP, 3D MP-RAGE, 3D VIBE) com diferentes parâmetros tendo os autores verificado que a sequência que proporcionou imagens de melhor qualidade da mandíbula foi o 3D VIBE. O ensaio das sequências foi feito num voluntário sendo o 3D VIBE depois utilizado para adquirir imagens em fantasmas (tubos de acrílico e osso seco da mandíbula). A partir das reconstruções 3D da mandíbula obtidas com RM 3D foram efetuadas algumas medidas e avaliada a sua exatidão, comparando com medidas obtidas a partir de imagens de TC (*considerado o gold standard*).

ou se seria necessário procurar outras alternativas. Após essas experiências, tendo em atenção os objetivos deste estudo, verificou-se que a qualidade das arcadas era suficiente para efetuar o co-registo destas, com os volumes obtidos em fonação, e ter uma ideia do espaço ocupado pelos dentes na cavidade oral. Se o objetivo fosse outro, provavelmente, seria necessário recorrer a outras estratégias. Os parâmetros que resultaram do processo de otimização da sequência VIBE (3D arcadas) para esta finalidade são apresentados, na secção seguinte, na tabela 5.3.

O trabalho realizado durante estas sessões permitiu-nos, para além disto, aferir e redimensionar o *corpus* final a adquirir na medida em que, estabelecido o protocolo de aquisição, nos é permitido ter uma ideia do tempo total de aquisição. Para lá do problema relacionado com as arcadas, as imagens resultantes desta fase de testes serviram para ir ensaiando várias alternativas de processamento de imagem.

5.3.3.4 Protocolo de aquisição de RM

Após a leitura do termo de consentimento informado e esclarecimento de questões relativas ao procedimento, os voluntários deram o seu consentimento para participação no estudo através do preenchimento do formulário desenvolvido e adaptado para aquisição de imagens de RM, no contexto de estudos de produção de fala (conforme consta no anexo C).

Antes da realização da sessão de aquisição de dados, todos os informantes responderam a um questionário de rastreio (obrigatório) relacionado com a realização do exame de RM, com recurso ao formulário utilizado na unidade de RM onde foram adquiridas as imagens (ver anexo D).

A aquisição de imagem foi efetuada com os falantes em posição de supinação, com os joelhos apoiados e fletidos para maior conforto. A cabeça e pescoço foram posicionados e imobilizados, nas antenas selecionadas, com almofadas de contenção para prevenir movimentos (rotação e/ou inclinação) durante a aquisição das imagens. Em todas as sessões de aquisição de dados os informantes utilizaram auscultadores adequados de forma a minimizar os índices de ruído e permitir comunicação com a zona de consola.

O plano médio-sagital do falante foi alinhado com o cursor luminoso vertical e a linha orbito-meatal (LOM) colocada perpendicular ao plano da mesa. A centragem foi efetuada na linha média sensivelmente a meio da área abrangida pelas duas antenas. Foi dada a oportunidade de utilizar o sistema de espelhos amovível, que faz parte da bobine de crânio, contudo nenhum dos falantes optou pela sua utilização.

Todos os falantes tiveram acesso prévio ao *corpus* a adquirir, para que pudessem efetuar algum treino, nomeadamente para controlo da respiração e sustentação do som durante o período de aquisição de cada sequência.

O protocolo teve início, como é habitual em qualquer estudo de RM, com uma sequência de localização rápida, onde foram efetuadas imagens de referência, nos vários planos, com o falante em repouso (sem produção de fala).

De seguida, foi programado um único corte no plano sagital, na linha média para obtenção

de informação 2D. A realização de um único corte permite a obtenção de uma imagem, com um tempo de aquisição relativamente curto e com boa resolução espacial. Foi utilizada uma sequência TSE (maior rapidez) com ponderação em T1, já otimizada anteriormente no âmbito de um estudo conduzido pela nossa equipa (Martins, 2007; Martins *et al.*, 2008). A utilização de um fator turbo (ETL) relativamente elevado pode conduzir a uma descaracterização da sequência no que diz respeito à ponderação obtida, o que, neste caso, é irrelevante, uma vez que não se pretende efetuar caracterização de lesões mas apenas a visualização do trato vocal e dos diferentes articuladores. Na tabela 5.2 encontram-se, de forma mais detalhada, os parâmetros utilizados nesta aquisição.

Depois da sequência ter sido lida e mantida em modo de espera foi comunicado ao falante (através do intercomunicador) qual seria o som que deveria sustentar e a palavra de referência para o eliciar. Nesse momento, foi também fornecida a indicação relativa ao tempo de aquisição da sequência. Por exemplo, para o /l/ em posição inicial, em contexto de [a], foi solicitado ao falante que produzisse o /l/ como na palavra **laca** e indicado o tempo de aquisição de 6 s. O falante foi instruído para sustentar o som durante o período em que ouvisse o barulho e para produzir a restante palavra, quando deixasse de o ouvir. Só depois de dadas as instruções é que a sequência foi lançada, começando a aquisição de imediato. A figura 5.1 ilustra todo este processo, repetido para cada um dos sons do *corpus* que foi adquirido.

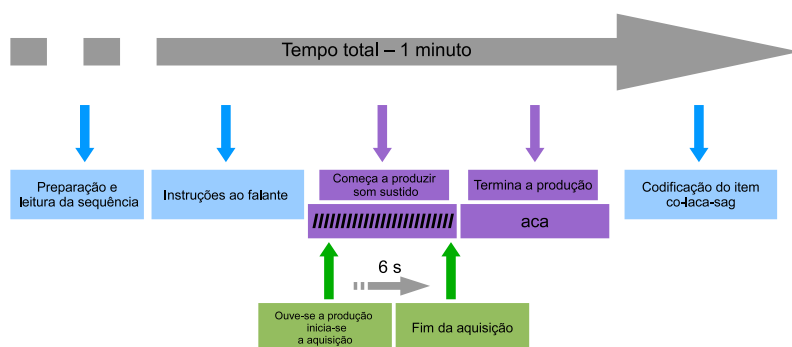


Figura 5.1: Sequência de passos para efetuar a aquisição das imagens. Exemplo para a eliciação do /l/, em início palavra, como em **laca**.

Após a aquisição da imagem 2D, procedeu-se à marcação e aquisição da sequência EG rápida (3D VIBE), também no plano sagital. Esta foi marcada de um lado ao outro da articulação temporo-mandibular do falante. Os parâmetros utilizados, que resultaram do processo de otimização, encontram-se especificados na tabela 5.3 (3D fonação). As instruções fornecidas foram as que atrás descrevemos, variando apenas a informação do tempo de aquisição (no caso foi de 19 s).

Sempre que houve dúvida quanto à produção de um determinado som, foi efetuada nova aquisição, tendo sido guardadas todas as repetições, para posterior análise. As produções foram monitorizadas através do intercomunicador (consola/sala magnete) ou mesmo com a presença do investigador dentro da sala, sempre que foi considerado necessário.

No final, após a recolha de todos os volumes com o falante em fonação, foi adquirido o volume de dados para posterior extração das arcadas dentárias, utilizando também a sequência 3D VIBE, com melhor resolução espacial, como já referido na secção anterior. O tempo de aquisição

Tabela 5.2: *Parâmetros de aquisição utilizados na sequência TSE T1 marcada, na linha média, no plano sagital.*

Parâmetro	Valor
<i>Sequência</i>	TSE
<i>Modo</i>	2d
<i>TR</i>	400 <i>ms</i>
<i>TE</i>	8 <i>ms</i>
<i>FA</i>	120
<i>ETL</i>	21
<i>FOV (leitura)</i>	240
<i>FOV (fase)</i>	240
<i>Intervalo entre cortes</i>	N.A.
<i>Orientação</i>	Sagital
<i>Número de cortes</i>	1
<i>Espessura de corte</i>	5 <i>mm</i>
<i>Resolução (leitura)</i>	256
<i>Resolução (fase)</i>	230
<i>Número de aquisições</i>	1
<i>Tempo de aquisição</i>	6 <i>s</i>
<i>Largura de banda</i>	283 <i>Hz/pixel</i>
<i>Dimensão pixel</i>	1.0 x 0.9 <i>mm</i>
<i>Espaçamento entre pixel</i>	0.9375

foi, aproximadamente, 3 minutos (tabela 5.3 - 3D Arcadas). Neste caso, o falante não produziu qualquer som. O tempo total que cada falante permaneceu dentro da *gantry* variou entre os 60 e os 75 minutos.

5.3.3.5 Recolha de dados acústicos

Posteriormente às sessões de aquisição de imagem, foi efetuada a recolha de dados acústicos para seis dos falantes que participaram nas aquisições de RM⁷. O objetivo desta recolha é possibilitar comparação entre os dados acústicos e os dados articulatórios obtidos através de RM, para os mesmos falantes. A gravação do sinal acústico foi efetuada na câmara insonorizada do Laboratório de Ciências da Fala da ESSUA na semana seguinte à recolha de dados de imagem.

Para manter um ambiente semelhante ao da RM, os falantes efetuaram a gravação numa posição deitada (em supinação), com um apoio sobre a cabeça (semelhante à posição da cabeça posicionada na antena de RM), com auscultadores nos ouvidos e apoio sob os joelhos.

Foi utilizado um microfone externo unidirecional (*AKG, C535EB II*), colocado num ângulo de 90° a cerca de 10 *cm* da boca do falante, e placa de som externa de 24 *bits* (*Cakewalk, UA-25 EX*) ligada a um PC. A frequência de amostragem usada foi de 22 *kHz*. Os segmentos gravados foram os utilizadas nas aquisições RM para eliciar o som em estudo. As palavras foram apresentados ao falante numa folha de papel, tendo o falante produzido a palavra com uma velocidade de elocução considerada normal e em versão sustida. Posteriormente, para segmentação

⁷Por motivos de ordem pessoal não foi possível efetuar gravação do sinal acústico para o falante AS.

Tabela 5.3: *Parâmetros de aquisição da sequência EG 3D FLASH VIBE otimizada para aquisição de imagens durante elocução dos diferentes fonemas (3D Fonação) e para as arcadas dentárias dos falantes (3D Arcadas).*

Parâmetro	3D Fonação	3D Arcadas
<i>Sequência</i>	<i>FLASH</i>	<i>FLASH</i>
<i>Modo</i>	3d	3d
<i>TR</i>	4.32 ms	20 ms
<i>TE</i>	1.57 ms	3.69 ms
<i>FA</i>	10°	12°
<i>ETL</i>	NA	NA
<i>FOV leitura</i>	270	256
<i>FOV fase</i>	237	256
<i>Sobre-amostragem (fase)</i>	10%	0%
<i>Sobre-amostragem (corte)</i>	0%	38.5%
<i>Distância</i>	20%	20%
<i>Orientação</i>	<i>Sagital</i>	<i>Axial</i>
<i>Número de cortes por slab</i>	52	104
<i>Espessura corte (ef.)</i>	2 mm	1 mm
<i>Resolução (base)</i>	256	256
<i>Resolução (fase)</i>	70 %	100%
<i>Resolução (corte)</i>	100 %	64%
<i>Número de aquisições</i>	1	1
<i>Imagem paralela</i>	<i>GRAPPA</i>	<i>GRAPPA</i>
<i>Fator de aceleração (PE)</i>	2	2
<i>Linhas referência (PE)</i>	24	24
<i>Assimetria ecos (modo)</i>	<i>fraco</i>	<i>fraco</i>
<i>Gradientes (modo)</i>	<i>normal</i>	<i>rápido</i>
<i>Tempo de aquisição</i>	19 s	3.29 min
<i>Largura Banda (BW)</i>	350 Hz/pixel	130 Hz/pixel
<i>Dimensão do voxel</i>	$1.5 \times 1.1 \times 2 \text{ mm}^3$	<i>isotrópica</i> (1 mm^3)
<i>Espaçamento entre pixel</i>	1.05 mm	1 mm

e anotação dos ficheiros áudio, recorreu-se ao programa *Praat* (Boersma e Weenink, 2009).

5.3.4 Processamento de Imagem

Para além do processo de otimização de sequências e da própria aquisição das imagens de RM, uma parte significativa deste trabalho está relacionada com a exploração de várias ferramentas de processamento de imagem e de métodos de segmentação para extração de dados articulatórios a partir do conjunto de imagens RM (2D e 3D) obtidas. Este trabalho visa, no imediato, a sua aplicação ao estudo dos sons laterais como nos propusemos fazer nesta etapa do nosso estudo, mas pretende-se que se venha a estender à análise de toda a base dados adquirida, que inclui outros sons do PE. Assim, uma vez que no capítulo 3 já se efetuou uma descrição dos métodos e das ferramentas de processamento de imagem a que se recorreu para as tarefas de segmentação, apresentaremos nas secções seguintes os passos envolvidos na segmentação das várias estruturas de interesse (e.g., trato vocal, língua - 2D e 3D).

5.3.4.1 Segmentação das arcadas e trato vocal

Obtenção de funções de área

A segmentação das vias aéreas com o intuito de obter a variação das áreas ao longo do trato vocal envolveu uma série de etapas e constituiu uma das tarefas de processamento e segmentação mais complexas deste trabalho.

A primeira etapa deste processo passou pela segmentação da mandíbula e da maxila. A segmentação destas estruturas foi efetuada a partir do volume de dados adquirido, especificamente com este intuito, para cada um dos falantes, como referido na secção anterior. A mandíbula e a maxila foram segmentadas, separadamente, obtendo-se duas máscaras binárias. A ferramenta utilizada para segmentação foi o *MeVisLab* com recurso a um módulo de segmentação que utiliza o *Live-wire*.

Uma vez que o *MeVisLab* não lê diretamente os ficheiros em formato *DICOM*, a primeira tarefa foi a importação das imagens através do módulo *DICOM Import* para um formato próprio do *MeVisLab*. Com a função *DICOM Browser* o volume de interesse foi carregado e visualizado através de um dos módulos de visualização disponíveis - (*VIEW 2D*). Este módulo está conetado ao módulo *CSO LiveWire processing* e, através deste, é efetuada a segmentação da estrutura corte a corte, se não for utilizada interpolação. Para tornar mais rápido o processo de segmentação recorremos à função (*CSO Interpolate*) que efetua interpolação entre os *CSO's* obtidos, se estes forem paralelos entre si.

De seguida, os *CSO's* são convertidos em imagens (*CSO ConvertToImage*) sendo criada uma máscara binária que, na fase seguinte deste processo, será co-registada com cada um dos volumes obtidos durante a produção dos sons constituintes do *corpus*. As máscaras, para cada uma das estruturas (mandíbula e maxila) foram gravadas como imagens e identificadas (e.g., *siglafalante-mandible*).

Na figura 5.2 pode observar-se um exemplo de um corte segmentado da mandíbula (esquerda) e da maxila (ao centro), para um dos falantes (JPM). Na imagem da direita apresenta-se uma malha superficial criada através do módulo *IsoSurface* e visualizada através da função *SoExaminerView*. Para não complicar a tarefa de co-registo foram efetuadas algumas simplificações: 1) o ramo da mandíbula não foi totalmente incluído nas segmentações; 2) nem sempre o palato duro foi completamente segmentado e 3) os dentes não foram delineados de forma individualizada.

Com estas opções não se obtiveram reconstruções perfeitas destas estruturas, como fica bem patente nas imagens apresentadas ao longo do trabalho. Todavia, assumiu-se esta posição porque constatou-se que ao incluir o ramo da mandíbula e o palato duro o custo computacional seria maior e o co-registo iria ser bastante mais difícil devido à forma mais complexa (e.g., curvatura do palato) e diferente orientação anatómica do corpo e do ramo da mandíbula. Tendo em atenção os nossos objetivos, uma segmentação mais perfeita destas estruturas não traria grande benefício uma vez que o palato duro é perfeitamente identificável nas imagens e a identificação das passagens laterais não depende da visualização do ramo da mandíbula.

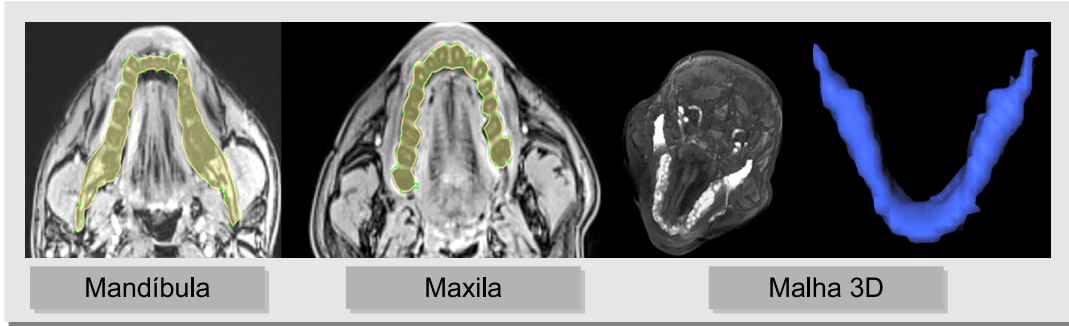


Figura 5.2: Exemplo de segmentação das arcadas dentárias do falante JPM - mandíbula (esquerda), maxilar (centro) e malha 3D da mandíbula (direita).

Para efetuar o co-registo destas máscaras, com cada um dos volumes obtidos em fonação, foi criada outra rede *MeVisLab*. A figura 5.3 ilustra esta etapa através da rede implementada para o efeito.

Primeiro, fez-se a importação das máscaras da mandíbula e da maxila e carregou-se o volume pretendido (zona identificada a tracejado na figura 5.3). Uma vez que as máscaras e o volume apresentam diferente resolução, o passo seguinte foi efetuar uma re-amostragem das máscaras e do volume para obter dimensão do *voxel* igual. Selecionou-se *voxel* isotrópico com dimensão de 1 mm. Para tal, recorreu-se a um filtro *Lanczos3*, dos muitos filtros de re-amostragem disponibilizados através do módulo *Resample 3D* e da opção *Resampling filters*.

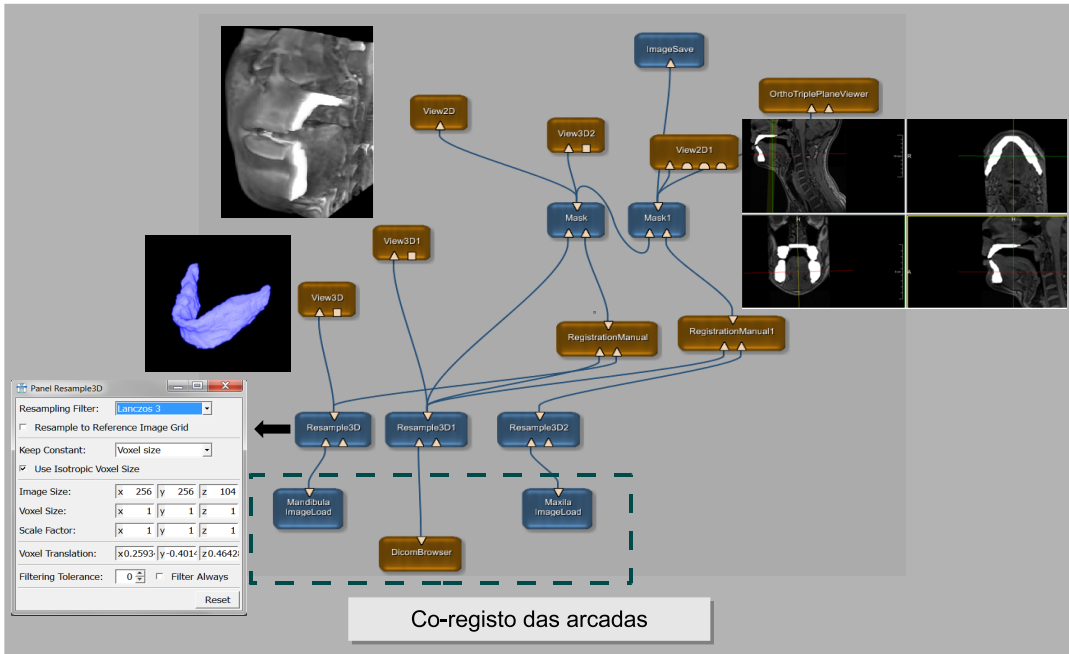


Figura 5.3: Rede *MeVisLab* criada para efetuar o co-registo das arcadas dentárias dos falantes com os volumes obtidos durante a produção de fonemas.

De seguida, de forma separada, procedeu-se ao registo da mandíbula e da maxila, de forma manual, através da funcionalidade *Registration Manual*. Esta função permite, através do

uso do rato e com visualização simultânea nos três planos ortogonais (com a opção *interactive affine 3D coregistration*), sobrepor e deslocar as máscaras com movimentos de rotação e translação até se obter um registo correto das máscaras (da mandíbula e da maxila) com o volume de interesse. O volume com as arcadas co-registadas foi guardado como imagem e identificado como (e.g., *lcr-sal-mand-max*). Esta codificação permite identificar o falante, o som e indica que o volume já tem incluídas as máscaras das arcadas do falante. Este processo foi repetido para cada um dos falantes e som adquiridos.

A etapa seguinte teve como objetivo obter uma reformatação do volume de forma a obter planos, o mais possível, perpendiculares à linha média do trato vocal. Tendo em atenção a forma do trato vocal e os módulos disponibilizados no *MeVisLab*, a função que nos pareceu mais indicada para obtenção dos cortes perpendiculares, a cada ponto, do trato vocal foi a Reconstrução Multiplanar Curva (*Curved Multiplanar Reformatation*). A figura 5.4 apresenta a rede utilizada para este efeito.

Assim, após o carregamento do volume relativo a um determinado som, já com as arcadas co-registadas, procurou-se uma imagem na linha média na qual foi definida uma via colocando várias sementes ao longo do trato vocal (função *path-definition*). Esta via foi marcada desde os lábios até à glote (neste falante coincidente com a vértebra cervical - C5). A espessura definida para a reformatação curva foi de 2 mm.

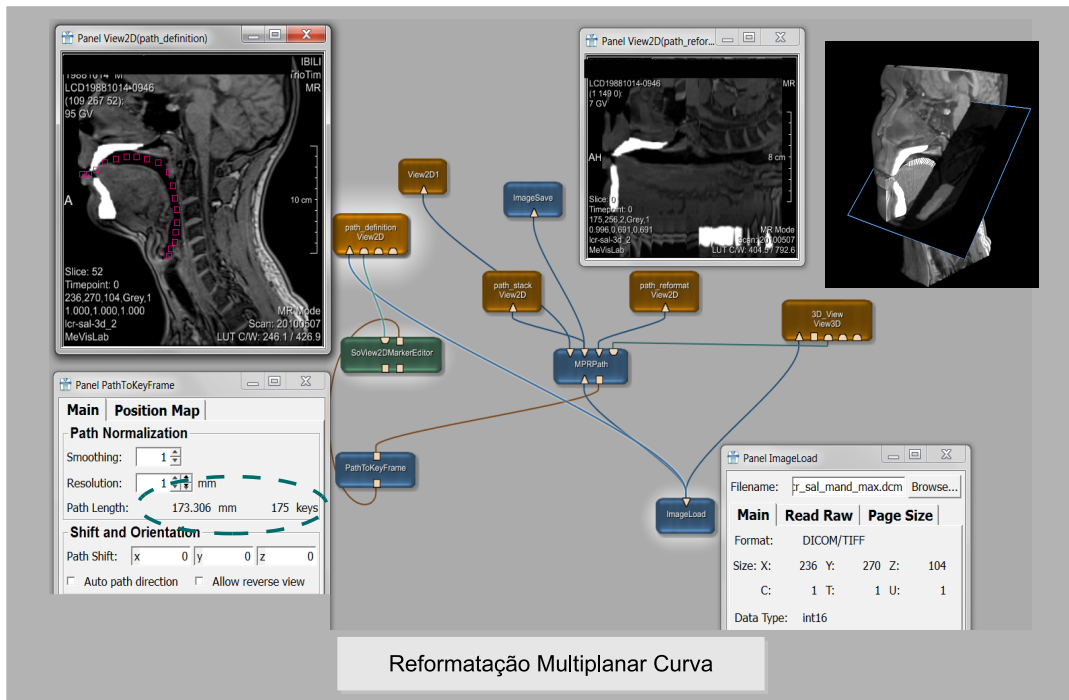


Figura 5.4: Rede *MeVisLab* utilizada para efetuar Reconstrução Multiplanar Curva ao longo do trato vocal. Exemplo para o falante LCR na produção de /l/ como em *sal*.

Efetuando a reconstrução obtiveram-se cortes, a cada milímetro, ao longo do trato vocal. O número de *frames* (cortes) é variável para cada falante (em função da dimensão do trato vocal) e para cada som, uma vez que a posição da glote pode estar mais ou menos elevada ou poder haver

arredondamento dos lábios. A via reformatada pode ser mostrada com a função *Path-reformat*.

Através da funcionalidade *Path-to-Key-Frame* obtém-se a informação relativa ao número de *frames* e ao comprimento (em *mm*) da via, anteriormente, definida. Este valor permite obter uma primeira aproximação (sobreestimada) relativa ao comprimento total do trato vocal. Posteriormente, através das segmentações, será possível obter um valor mais exato do comprimento do trato. Neste falante (LCR), para o /l/ produzido na palavra **sal**, o comprimento da via é de 173.3 *mm* e o número de cortes é 175. Na figura 5.4, no canto superior direito, pode observar-se um *rendering* do volume e a orientação dos vários cortes ao longo do trato. A orientação de um dos 175 cortes é também apresentada; esta corresponde a um corte oblíquo, perpendicular à linha média do trato vocal, na região uvular. Os resultados para cada som foram guardados e codificados da seguinte forma: *lcr-sal-mpr*.

Só neste momento se reuniram as condições para efetuar a segmentação do trato vocal a partir dos cortes reconstruídos pelo processo acima descrito. A rede *MeVis* utilizada é semelhante à que já foi descrita para a segmentação das arcadas dentárias utilizando como técnica de segmentação o *Live-wire*. Sucintamente, o volume foi carregado e visualizado através do módulo *VIEW 2D* e iniciou-se a segmentação. Para o exemplo que utilizamos acima, para o falante LCR, a reformatação curva deu origem a 175 cortes, espaçados de 1 *mm*. A segmentação de todos os cortes resultaria num esforço de segmentação e computacional muito elevados. Assim, foi definido que da glote até à zona do véu palatino se segmentaria uma imagem de cinco em cinco (com início na glote) o que corresponde a obter informação a cada 5 *mm* (aproximadamente) e, na cavidade oral, se reduziria o espaçamento para 3 cortes uma vez que é nesta zona que se necessita de maior detalhe para poder obter as áreas das passagens laterais permitindo determinar, com mais precisão, a sua extensão. Não obstante esta opção tomada, a qualquer altura se poderão segmentar, as mesmas imagens, com diferente resolução uma vez que a aquisição volumétrica efetuada assim o permite. Numa fase posterior, para efetuar modelação, fizemos segmentação de 3 em 3 cortes, ao longo de todo o trato vocal, para alguns dos sons do *corpus*.

Os contornos obtidos (*CSO's*) foram guardados através da função (*CSOsave*) e, posteriormente, podem ser exportados para outros programas. Depois de guardados, estes contornos podem ser novamente carregados, a qualquer altura, importando o volume de interesse e os *CSO's* recorrendo à funcionalidade *CSOLoad*.

Para efetuar a leitura dos *CSO's* e para extração de dados foi utilizado o *software MATLAB* (*MATLAB 7.5.0, The Mathworks Inc., Natick, MA, 2007*). O trabalho de criação dos códigos em *MATLAB*, para obter a informação, foi efetuado por um dos orientadores deste trabalho⁸ e acompanhado pela autora. Desta forma, obtiveram-se: sobreposição dos contornos em 2D, áreas e funções de área do trato vocal, comprimento do trato vocal, estimativa da extensão e área das passagens laterais, extensão do contacto alveolar, etc. Foram ainda criadas visualizações 3D dos cortes segmentados (batizadas como “lagostas”).

⁸Professor Doutor António Joaquim da Silva Teixeira.

5.3.4.2 Segmentação da língua

Obtenção de contornos e processamento 2D

A segmentação das imagens 2D foi exclusivamente efetuada recorrendo ao *MeVisLab* utilizando o módulo de segmentação baseado no *Live-wire*. A rede utilizada para obter a segmentação da língua em 2D apresenta-se na figura 5.5.

O processo de segmentação é simples e relativamente rápido (apenas uma imagem). Os *CSO's* foram exportados para *MATLAB* e, através deste, foram obtidas imagens com sobreposição dos contornos para as várias produções. A comparação de contornos através da sobreposição de contornos permite uma análise meramente qualitativa. Permite, por exemplo, identificar as zonas da língua em que há mais variação entre os contornos obtidos para diferentes produções.

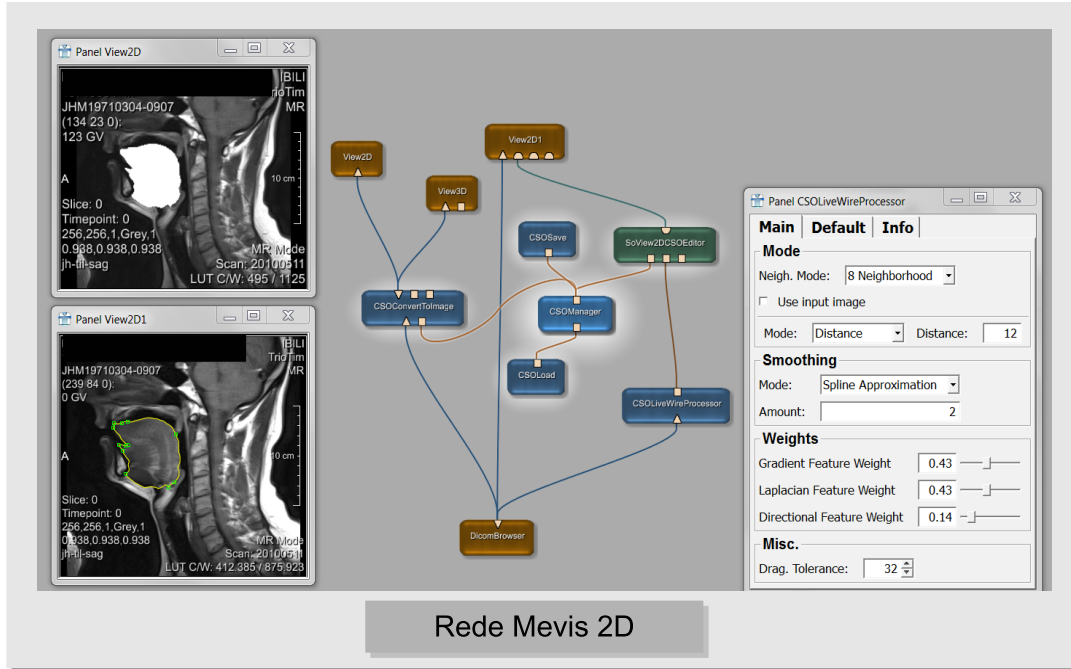


Figura 5.5: Rede *MeVisLab*, com módulo *Live-wire*, para segmentação das imagens 2D da língua.

Para que fosse possível quantificar as diferenças entre os contornos da língua, para as diferentes posições e contextos vocálicos considerados, recorreu-se a uma métrica - a figura de mérito de *Pratt*, também conhecida como Índice de *Pratt* (IP), proposta por Pratt em 1978. Em trabalhos anteriores, para estudo de efeitos coarticulatórios, já havíamos recorrido a esta métrica para estabelecer a comparação de contornos (veja-se, por exemplo, Martins *et al.*, 2008).

O Índice de *Pratt* é definido por:

$$IP = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 + \alpha d_i^2} \quad (5.1)$$

em que N é o número de pontos com correspondência entre os contornos, d_i é a distância euclidiana

entre dois pontos correspondentes e α é uma constante relacionada com a dimensão do contorno.

Tendo por base o valor proposto por *Pratt*, seguido noutras referências consultadas (e.g., Silva, 2005) considerou-se $\alpha = 1/9$. Para obter correspondência entre os vários pontos do contorno, seleciona-se o contorno com o menor número de pontos. Para cada ponto desse contorno, o ponto mais próximo do contorno a comparar é considerado o ponto correspondente. A figura de mérito de Pratt, fornece uma medida global do comportamento das distâncias entre os contornos. Para o valor de IP assim calculado concorrem todas as diferenças (regionais) entre o par de contornos. Não é possível afirmar, por exemplo, em que zona da língua as diferenças foram maiores.

O IP é uma medida relativa que devolve valores dentro do intervalo]0–1]. Se dois contornos tiverem uma correspondência perfeita (iguais) o IP terá o valor de 1. A utilização de IP parciais, considerando separadamente diferentes zonas da língua, poderá vir a revelar-se mais atrativa neste contexto (veja-se Silva *et al.*, 2013b).

Processamento dos dados (3D) - Obtenção de volumes e visualizações

A segmentação da língua em 3D foi efetuada com duas abordagens distintas. Uma utilizando o *Live-wire* através da implementação no *MeVisLab* e a outra com recurso a contornos ativos no *ITK-SNAP*.

A rede *MeVisLab* utilizada não varia muito das que já descrevemos para a segmentação do trato vocal. Neste caso foi necessário efetuar a segmentação de uma série de cortes (todos onde a língua é visualizada) ao longo do volume (cerca de 22 cortes mas varia com os falantes). O plano selecionado para efetuar as segmentações foi o plano sagital. É possível obter malhas 3D da língua e observá-las dentro do volume onde foram co-registadas as arcadas possibilitando visualizações 3D que ajudam a perceber a forma como a língua se posiciona dentro da cavidade oral condicionando a forma como o ar se distribui ao seu redor. A figura 5.6 mostra a rede *MeVisLab* utilizada para este efeito. Apesar de se utilizar um algoritmo semi-automático o processo não é muito eficiente.

Segmentação semi-automática da língua efetuada com o ITK-SNAP

Para além do processo de segmentação descrito acima, baseado no *Live-wire*, foi ainda testado um outro método de segmentação implementado no *ITK-SNAP*.

Descrevemos, de forma sequencial, todo o processo de segmentação da língua utilizando o *ITK-SNAP*. Como referência para a descrição e, na altura, para nos familiarizarmos com o algoritmo recorremos ao manual de utilização do *ITK-SNAP*⁹ e ao protocolo desenvolvido por Nguyen (2006) para segmentação dos ventrículos laterais, utilizando a mesma ferramenta.

O processo que se apresenta resultou de um trabalho de otimização, com realização de vários ensaios e ajustes nos parâmetros nas diferentes fases da segmentação, até se conseguirem os resultados pretendidos. Salientamos que, geralmente, os parâmetros (e.g., valores *threshold*) têm que sofrer ajustes ligeiros para cada um dos falantes mantendo-se, aproximadamente, constantes

⁹Disponível online em <http://www.itksnap.org>.

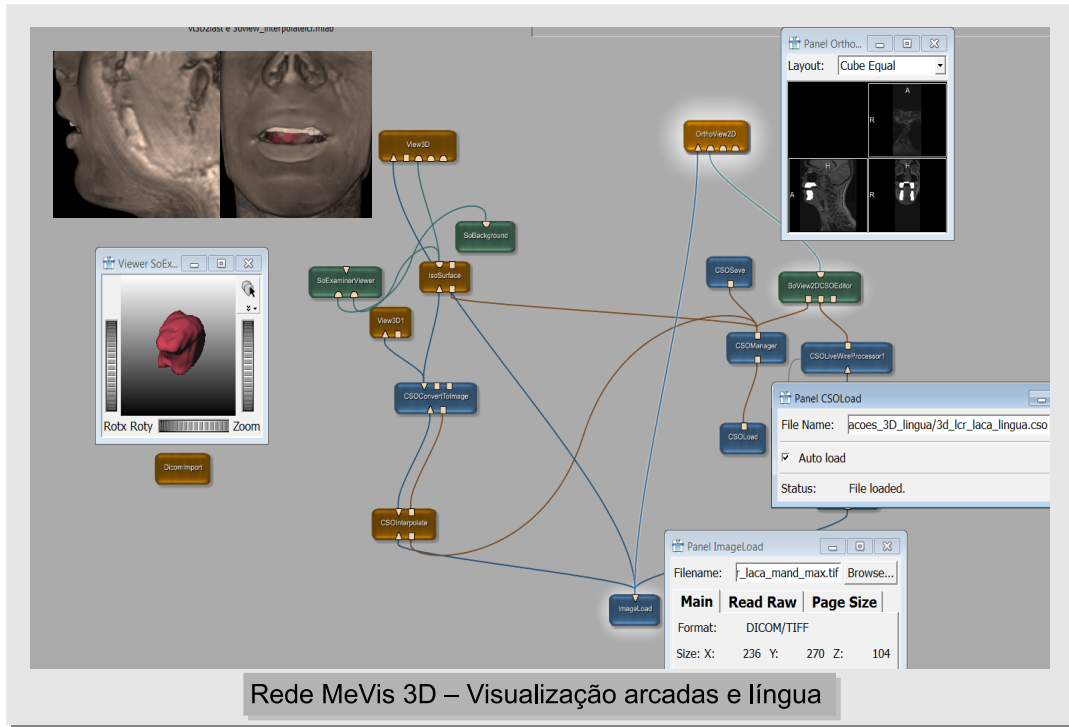


Figura 5.6: Rede *MeVisLab*, com módulo *Live-wire*, para segmentação e visualização das imagens 3D da língua.

nos diferentes volumes, para cada falante.

A primeira tarefa passa por carregar o volume que se pretende processar utilizando a função *Load greyscale image*, o formato de importação selecionado foi, neste caso, o formato *DICOM file*. A janela de visualização do *ITK-SNAP* apresenta o volume nos três planos ortogonais (sagital, coronal e axial) e tem ainda um espaço onde é possível efetuar a atualização da malha superficial (*mesh*) e visualizá-la a qualquer altura do processo de segmentação. A figura 5.7 apresenta a interface gráfica do *ITK-SNAP* com um exemplo de segmentação para o véu palatino. É possível gerir este espaço de acordo com a preferência do operador, no entanto, é desejável colocar as imagens com a orientação anatómica.

O *ITK-SNAP* permite efetuar delineação manual das estruturas ou segmentação automática. Selecionando a opção automática, imediatamente se abre um painel que guia o processo de segmentação. Para que a qualidade da segmentação seja otimizada e para minimizar o tempo e esforço de segmentação é necessário delimitar a região de interesse através da função *reset region*. Assim, foi definido o *Volume of interest* (VOI), neste caso, para a língua, guiada nos três planos, para evitar cortes na estrutura alvo (ver figura 5.8, à esquerda).

De seguida, procedeu-se a uma re-amostragem do volume de interesse selecionando a função *Resampling* que permite escolher um de entre quatro filtros. Como as imagens disponíveis não são isotrópicas ($1.059 \times 1.059 \times 2.0 \text{ mm}^3$) este passo é importante para obter *voxel* quadrado e melhorar os resultados da segmentação. Optou-se pelo *Windowed Sinc interpolation* que não sendo o mais rápido proporciona uma melhor qualidade (Nguyen, 2006).

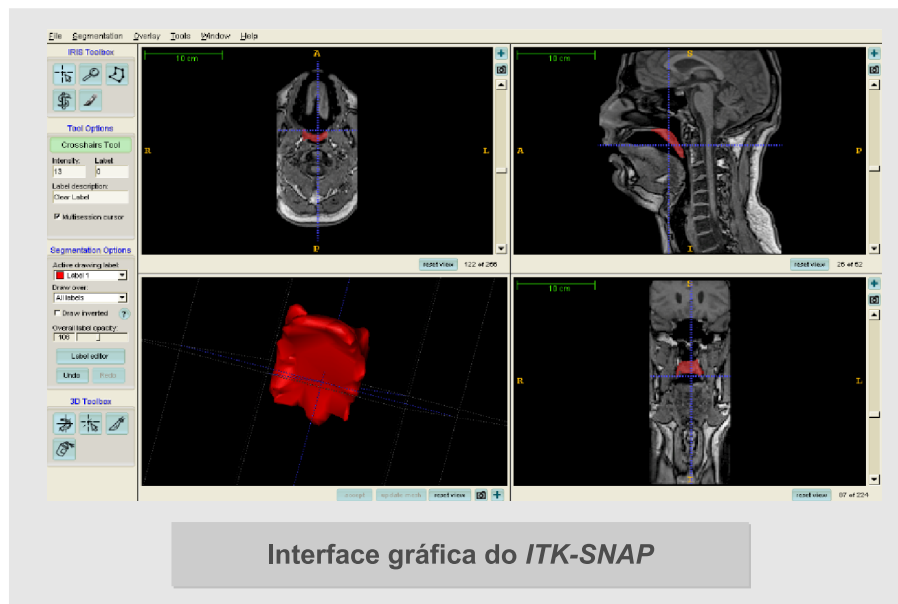


Figura 5.7: Interface gráfica do ITK-SNAP. *Display* do volume nos planos axial e sagital (em cima) e coronal (em baixo). Exemplo de segmentação numa produção do falante JPM. Malha superficial do véu palatino (em baixo, à esquerda).

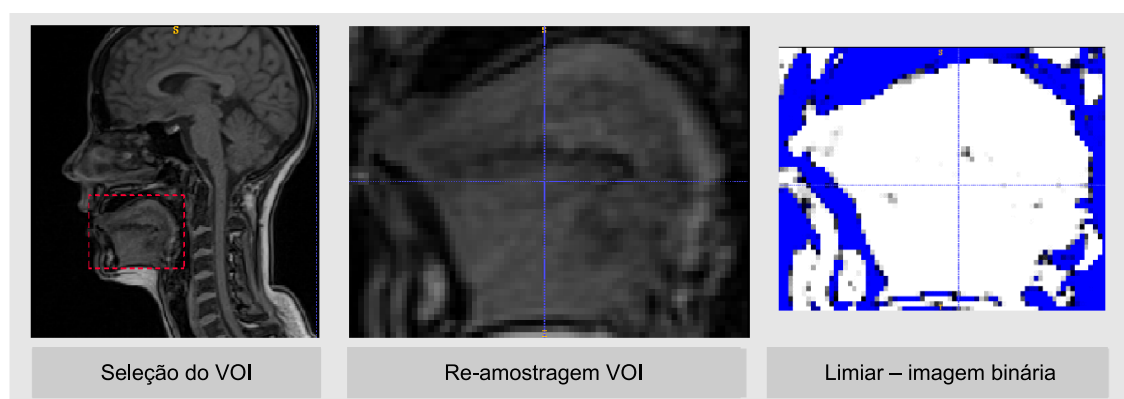


Figura 5.8: Seleção da região de interesse e re-amostragem dos dados.

Definida a zona de interesse e após re-amostragem dos dados inicia-se a fase de pré-processamento (*pré-processing step*) que visa preparar o volume para a segmentação.

Primeiro, tem que se optar por um de dois tipos de informação, presente na imagem, que guiará a evolução do contorno (*intensity region* ou *image edges*). A primeira alternativa implementa o algoritmo baseado em regiões de competição (*region competition snakes*) e funciona bem quando a estrutura a segmentar apresenta uma gama de intensidades bem definida, em relação ao fundo. A segunda opção implementa os contornos ativos geodésicos que permitem melhores resultados quando as estruturas apresentam um contorno com intensidade de sinal elevada (veja-se Yushkevich *et al.*, 2006 para mais detalhes).

Foram testadas as duas opções tendo recaído a escolha sobre a evolução do contorno governada por regiões de competição (*intensity region*) uma vez que os contornos da língua nas imagens não apresentam uma intensidade de sinal elevado. Por outro lado, como pelo menos a parte superior da língua e a sua parte posterior estão rodeadas por ar, a diferença de intensidade do objeto em relação ao fundo é considerável obtendo-se bons resultados.

Os maiores problemas surgem na parte inferior e lateral da língua uma vez que está rodeada por estruturas, também elas musculares, que apresentam nas imagens de RM uma intensidade de sinal aproximada à da língua. Nesse aspeto, uma delimitação precisa do VOI é extremamente importante, minimizando a necessidade de efetuar correções manuais.

Depois, têm que ser definidos os limiares superior e inferior de intensidade (com base na similaridade dos valores de intensidade das imagens) sendo semelhante a uma segmentação por *thresholding*. Escolhe-se também o grau de suavização a utilizar (ver figura 5.9). Desta forma, é criada uma imagem binária em que a estrutura de interesse (a língua) apresenta uma probabilidade de 1 e o fundo (ar ao seu redor) uma probabilidade de 0 (ver figura 5.8, à direita). De seguida, são colocados *3D balloons ou bubbles* dentro da estrutura, neste caso a língua. As *bubbles* são os chamados modelos deformáveis que se ajustam à região de interesse e às suas fronteiras (Fernandes, 2008; Yushkevich *et al.*, 2006).

Antes de iniciar o processo evolutivo têm que se selecionar os parâmetros e as forças internas e externas que vão governar a evolução do contorno. Com o algoritmo de regiões de competição, o termo da constante de propagação (α) é modulado pelas forças da imagem que, localmente, alteram a direção e força da propagação (a *snake* contrai-se quando a fronteira inclui parte do fundo e expande-se quando a fronteira está dentro da região requerida). Este algoritmo faz com que exista uma competição entre o objeto e o fundo (daí o seu nome) atingindo um estado de equilíbrio na convergência (Ho *et al.*, 2003). Este processo pode ser efetuado de forma intuitiva contudo, está disponível a formulação matemática subjacente.

A figura 5.10 ilustra este passo na segmentação. Para o α foi selecionado o valor de 1 e para γ o valor de 0.6 seguindo os valores referidos em Yushkevich *et al.* (2006) e Nguyen (2006). As forças de imagem são ainda equilibradas com a introdução de constrangimentos de fronteira (*boundary smoothness*).

Depois de algumas experiências verificámos que a colocação de várias bolhas de pequeno

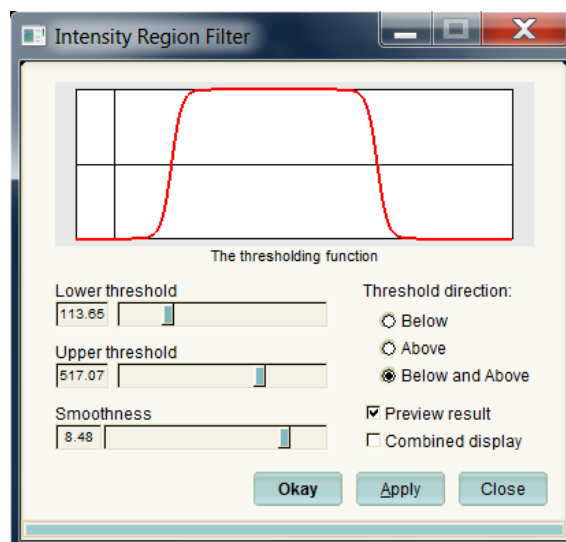


Figura 5.9: Limiares selecionados para criar mapa de probabilidades.

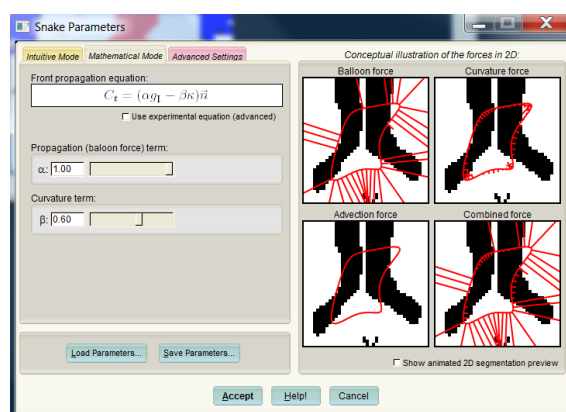


Figura 5.10: Seleção dos parâmetros para os termos da equação que governam o processo de evolução da *snake*.

raio, totalmente dentro da estrutura, permitia uma convergência e evolução mais rápida. A evolução é bastante rápida (alguns segundos) convergindo ao fim de, aproximadamente, 160 a 200 iterações. A figura 5.11 ilustra a colocação das bolhas dentro da língua e o processo de evolução através de várias iterações. Por vezes, houve colapso das bolhas havendo necessidade de repetir o processo.

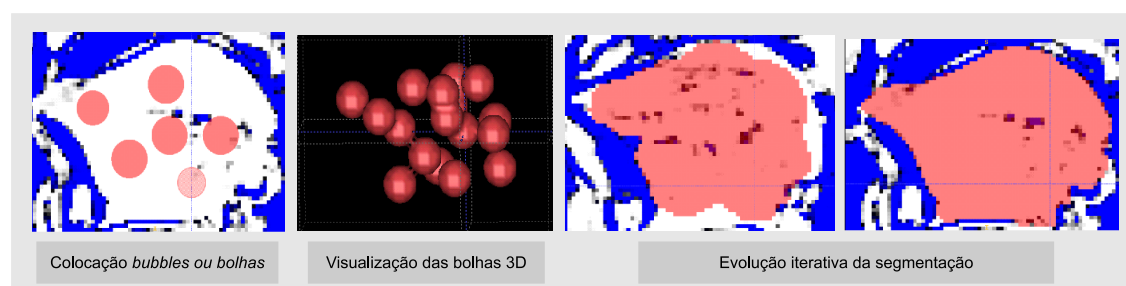


Figura 5.11: Colocação das bolhas (*balloons*) dentro da região de interesse e evolução iterativa da segmentação.

Terminado o processo (pelo operador) é possível obter uma malha superficial triangular que no *ITK-SNAP* é gerada com recurso a *Marching Cubes* (método proposto por Lorensen e Cline em 1987). Nesta fase, há necessidade de rever a segmentação proposta pelo algoritmo e efetuar correções utilizando as ferramentas disponíveis para segmentação manual. Assim, é possível adicionar ou remover *voxels* em zonas onde ocorreu extravasamento para zonas fora do objeto de interesse ou não o incluiu na sua totalidade. No nosso caso esta tarefa demorou alguns minutos. A figura 5.12 apresenta um exemplo de zonas onde foi necessário efetuar correção manual e o resultado final da malha (direita).

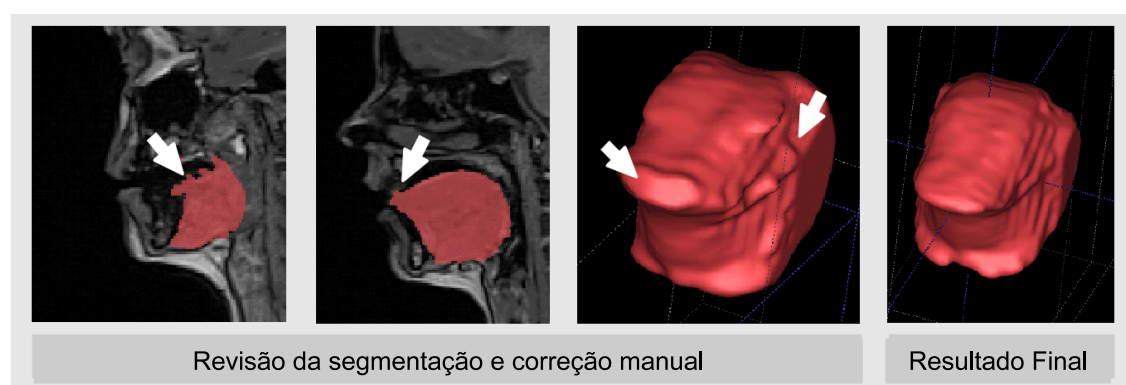


Figura 5.12: Revisão das segmentações e correções manuais efetuadas pelo operador.

Podem ainda ser aplicadas sobre a malha algoritmos de suavização e decimação. A figura 5.13 mostra o efeito de diferentes níveis de suavização aplicados sobre as malhas (três imagens da esquerda, respetivamente com 0.5, 1.0, 1.5 e 2.0 e de decimação (imagem da direita). Tendo em atenção que um nível de suavização muito elevado pode “apagar” aspetos importantes como a presença de *groove*, optou-se por utilizar um nível de suavização de 1.0 nas malhas obtidas para todas as segmentações.

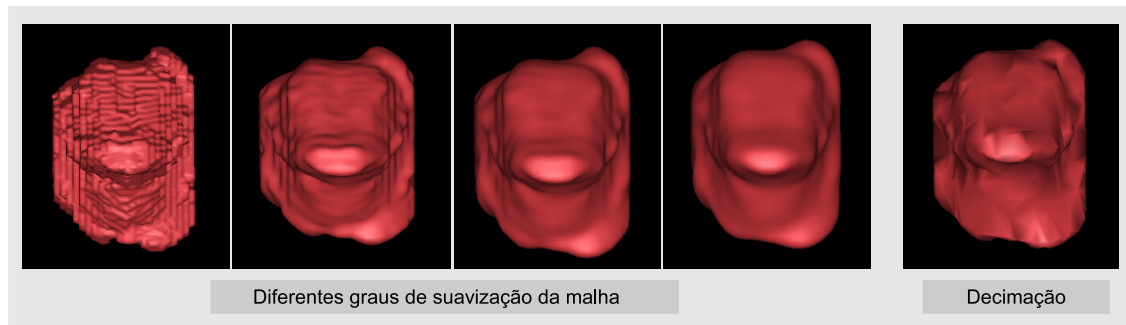


Figura 5.13: Influência dos algoritmos de suavização e decimação nas malhas superficiais.

O resultado da segmentação propriamente dita e a malha podem ser guardados em vários formatos (e.g., *.gipl* ou *.stl* para as malhas) e exportadas para serem utilizadas com outros programas. As estatísticas relativamente ao volume podem ser observadas e guardadas através da função - *Volume Statistics*. A figura 5.14 apresenta o ficheiro devolvido pelo *ITK-SNAP*, para um dos volumes, após segmentação.

```
#####
# SNAP Voxel Count File
# File format:
# LABEL: ID / NUMBER / VOLUME / MEAN / SD
# Fields:
# LABEL      Label description
# ID          The numerical id of the label
# NUMBER      Number of voxels that have that label
# VOLUME      Volume of those voxels in cubic mm
# MEAN        Mean intensity of those voxels
# SD          Standard deviation of those voxels
#####
Label 1      :      1 /      36664 /      81567.5
/      227.111 /      53.4004
```

Figura 5.14: Dados estatísticos relativos ao volume, fornecidos pelo *ITK-SNAP*.

5.3.4.3 Outras experiências de segmentação efetuadas

Encorajados pelos bons resultados obtidos com o *ITK-SNAP*, para as segmentações da língua, efetuaram-se algumas experiências de segmentação para outras estruturas, nomeadamente: véu palatino, trato vocal, arcadas dentárias e lábios. O processo de segmentação segue os passos já descritos para a língua havendo necessidade de definir as regiões de interesse (em função da estrutura) e alterar os parâmetros (e.g., limiares de intensidade) que serão diferentes dos utilizados para a língua.

No que diz respeito ao trato vocal, verificou-se a necessidade de selecionar duas regiões de interesse: uma para a zona da faringe, outra para a zona da cavidade oral. O método semi-automático funcionou relativamente bem para sons como as vogais ou para segmentação da zona da faringe permitindo uma segmentação rápida. A figura 5.15 mostra um exemplo de malha 3D do trato vocal a partir de segmentação semi-automática com algoritmo de regiões de competição, para uma vogal oral [u], referente a uma produção da falante MC. No caso de vogais houve apenas

necessidade de pequenos ajustes manuais podendo ser uma opção a considerar, no futuro, para processamento de imagens relativas a estes sons.

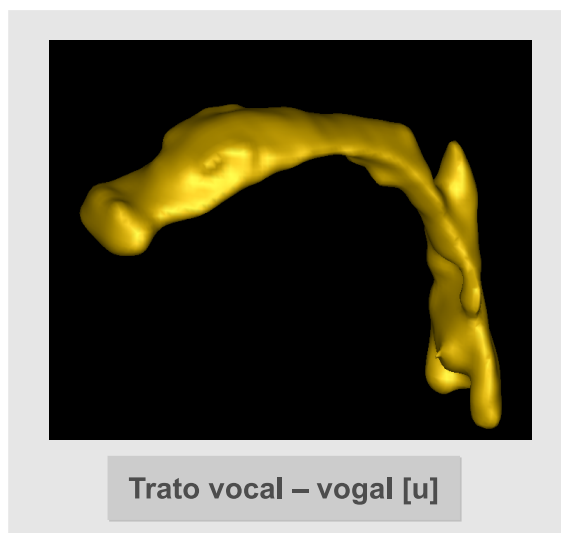


Figura 5.15: Exemplo de malha 3D resultante de segmentação do trato vocal para uma vogal, efetuada com algoritmo semi-automático no *ITK-SNAP*. Falante MC - vogal [u].

No entanto, para as consoantes laterais, uma vez que há descontinuidades (o fluxo de ar passa lateralmente e por cima da língua), o método não se mostrou muito eficiente. Ou seja, é possível fazer a segmentação recorrendo ao algoritmo de regiões de competição mas são necessários tantos ajustes manuais que o tempo de segmentação é pouco menor do que o de uma segmentação puramente manual. Ainda assim, foram efetuadas segmentações do trato com este método, para algumas produções, uma vez que nos proporcionam uma visualização 3D conjunta da língua e do trato vocal que se revelou interessante para ilustrar as passagens laterais e contactos da língua na zona alveolar (centralmente) e com o palato (lateralmente).

A figura 5.16 apresenta um exemplo de malhas 3D da língua e do trato vocal obtidas para três falantes (CO, JPM e LCR), para produções do /l/ como em [lak6].

Para o véu palatino o algoritmo semi-automático deu origem a bons resultados como se pode observar na imagem da direita (malha 3D do véu palatino - cor azul) da figura 5.17. Note-se contudo, que neste caso o véu palatino estava numa posição subida e afastado da língua permitindo delimitar um VOI que não incluísse o dorso da língua o que, a verificar-se, conduziria à progressão do algoritmo para esta estrutura. Assim, no caso de sons nasais (véu palatino baixo) seria provavelmente necessário dispendir mais tempo com correções manuais para obter a segmentação final.

No que concerne às arcadas dentárias apenas a segmentação manual se afigurou como opção utilizando o *ITK-SNAP*. A figura 5.17 apresenta um exemplo de malhas 3D, obtidas após segmentação, das arcadas dentárias da falante MC. À esquerda malha 3D da mandíbula e à direita malha referente à maxila. Como se pode observar, contrariamente ao que se conseguiu com o *Live-Wire* usando o *MeVisLab*, é possível obter uma segmentação que inclui o ramo da mandíbula e que

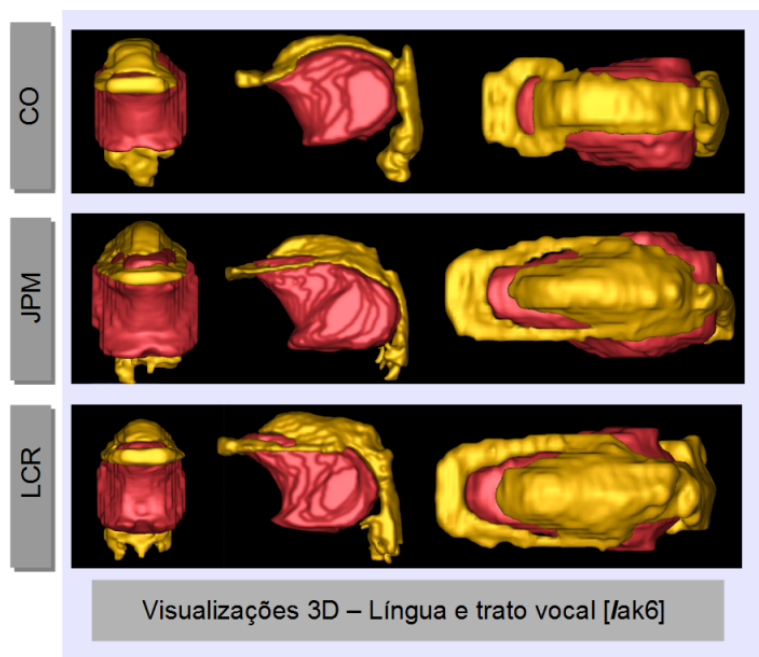


Figura 5.16: Malhas 3D da língua e trato vocal obtidas com o *ITK-SNAP*, para 3 falantes.

poderemos considerar mais perfeita. No entanto, a segmentação exclusivamente manual efetuada no *ITK-SNAP* torna-se extremamente demorada e cansativa para o operador. Poderia ser uma opção se houvesse apenas um falante, o que não é o caso no presente estudo.

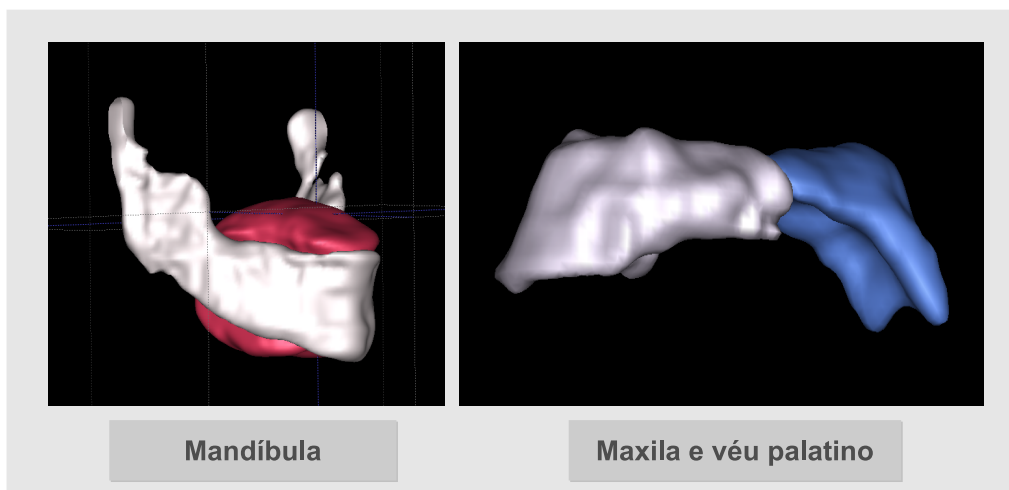


Figura 5.17: Exemplo de segmentação das arcadas dentárias e véu palatino da falante MC. À esquerda imagem da mandíbula e à direita imagem da maxila e véu palatino.

Finalmente, foi testado o método de *Region Growing* 3D neste caso através de uma implementação no *MeVisLab*. Este método é bastante eficiente, por exemplo, para segmentação da zona da faringe. Permite de forma muito rápida obter dados relativamente ao volume desta cavidade. De facto, após definição da região de interesse (ver imagem superior da figura 5.18)

selecionou-se uma vizinhança de *6-pixels* em 3D e foi necessário colocar apenas uma semente, na região selecionada, para obter a segmentação da cavidade faríngea. Na malha 3D apresentada na parte inferior da figura 5.18 (vista lateral e posterior da cavidade faríngea) não foi efetuada qualquer correção manual.

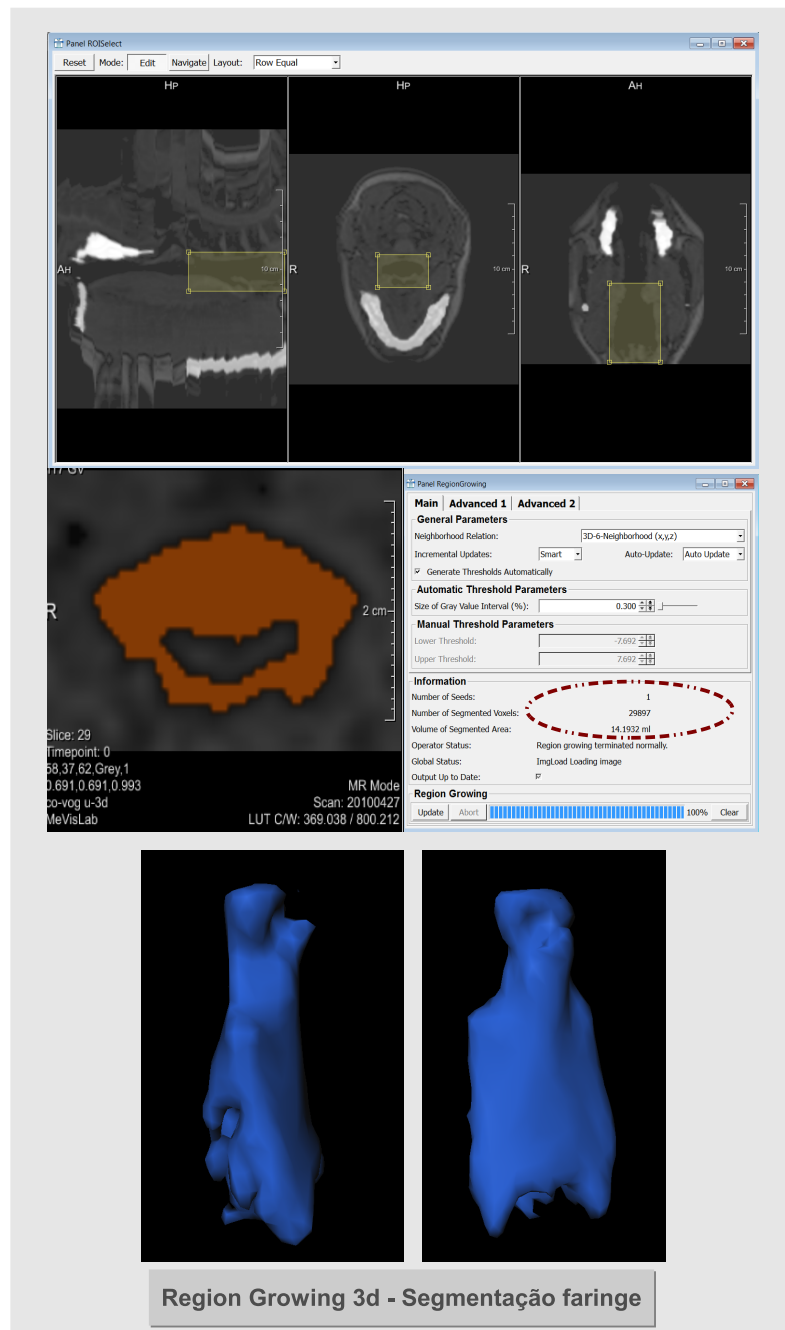


Figura 5.18: Exemplo de segmentação da faringe para uma vogal com recurso ao algoritmo de *Region Growing* 3D através do *MeVisLab*.

Na zona da cavidade oral, como já se antecipava, o método não funcionou bem não constituindo opção para tratamento dos dados relativos às consoantes laterais. Neste caso, teria

que se conjugar com outros métodos de segmentação (e.g., *live-wire*) ou utilizar filtros adequados (veja-se Wismueller *et al.*, 2008) que evitassem a progressão para lá dos lábios e, lateralmente, para a zona dos dentes.

O RG 3D poderá constituir um método alternativo para segmentação da língua ou para segmentação da faringe quando seja apenas essa a área de interesse como, por exemplo, comparar o volume da cavidade faríngea entre vogais orais/nasais ou entre fricativas vozeadas e não vozeadas (Proctor *et al.*, 2010b) ou, ainda, em situações que interesse quantificar o volume da coluna de ar nesta zona (e.g., doentes com apneia do sono).

Na tabela 5.4 apresenta-se o resumo do trabalho de processamento efetuado, onde se inclui informação sobre as estruturas segmentadas, as ferramentas utilizadas, o método de segmentação, o tempo despendido na segmentação (quando contabilizado), avaliação da qualidade (subjativa) e um pequeno comentário (relativo à eficiência ou às estruturas em que pode ser utilizado). O *score*¹⁰ utilizado para a avaliação da qualidade foi definido pela autora, considerando essencialmente se os resultados que se obtiveram responderam ao que era pretendido.

Tabela 5.4: *Resumo do trabalho de processamento efetuado.*

Estrutura	Ferramenta	Método Segmentação	Tempo	Qualidade	Comentários
Arcadas	SNAP MeVis	Manual <i>Live-wire</i>	quase 1 hora 30 minutos	+++ ++	Cansativo Pouco eficiente
Língua 2D	MeVis	<i>Live-wire</i>	20 s	++++	Simples
Língua 3D	MeVis SNAP SNAP	<i>Live-wire</i> <i>Snakes</i> Manual	10 minutos 6-7 minutos (ccm) Média de 30 minutos [22 a 41 minutos]	++ ++++ ++++	Eficiente Pouco eficiente, cansativo
Trato vocal 3D	Mevis MeVis SNAP	<i>Live-wire</i> <i>Region Growing</i> 3D <i>Snakes</i>	difícil contabilizar 2 minutos (ccm) 7-8 minutos (ccm)	++++ +++ +++ -	Pouco eficiente alternativa laterais eficiente (para faringe) adequado para vogais para as laterais
Véu palatino 3D	SNAP	<i>Snakes</i>	2 minutos (ccm)	++++	Potencial nas vogais nasais

5.4 Avaliação (preliminar) das segmentações efetuadas para a língua com *ITK-SNAP*

O estudo de avaliação efetuado e que se descreve nesta secção faz parte da publicação:

Martins, P., Oliveira, C., Silva, S., Silva, A., Teixeira, A. (2011). *Tongue segmentation from MRI images using ITK-SNAP: Preliminary Evaluation*. Proceedings of the IADIS CGVCVIP-2011, pages 3-10, Roma, Itália.

A validação dos métodos de segmentação utilizados não fazia parte dos objetivos deste trabalho. No entanto, a determinada altura do seu desenvolvimento, no que diz respeito às segmentações da língua efetuadas com o *ITK-SNAP*, pareceu-nos importante efetuar uma avaliação do desempenho do algoritmo. Optou-se por não lhe chamar um estudo de validação uma vez que não foi possível incluir e comparar segmentações efetuadas por vários observadores. No entanto, seguiram-se os mesmos passos e recorreu-se às métricas que, usualmente, se utilizam no

¹⁰Insuficiente (-) Razoável (++), Bom (+++) e Muito Bom (++++). Na tabela, o tempo total despendido na segmentação inclui as correções manuais (ccm).

contexto de estudos de validação. A questão fulcral era perceber se um método de segmentação semi-automático (baseado em regiões de competição) bastante eficiente (6-7 min) na delineação da língua, poderia substituir a segmentação manual, que é um método muito mais demorado e extremamente cansativo para o operador. Era também importante ter uma ideia se as visualizações 3D ou as malhas obtidas, a partir das segmentações, representavam “adequadamente” as verdadeiras características do objeto que representam. A base de dados que atualmente dispomos é muito vasta, como tal, a eficiência do método de segmentação torna-se extremamente importante. No entanto, o método deverá ser também reproduzível e exato. Se assim for, neste caso concreto, proporcionar-se-á estender os estudos de coarticulação em 2D já efetuados (Martins *et al.*, 2011a; Teixeira *et al.*, 2011) agora com informação 3D (e.g., estabelecer comparações da configuração da língua em 3D). Esta abordagem será interessante e útil, não só, para avaliação de efeitos contextuais para as consoantes laterais mas também para outros sons (e.g., oclusivas, fricativas).

No que diz respeito ao *ITK-SNAP* encontramos dois estudos que efetuaram testes de validação ao algoritmo, ambos relativos à segmentação do núcleo caudado e dos ventrículos laterais (Ho *et al.*, 2003; Yushkevich *et al.*, 2006), a partir de imagens de RM. Os resultados apresentados foram muito bons para essas estruturas (elevada eficiência e precisão) o que nos motivou a tentar averiguar qual o desempenho do algoritmo na segmentação de uma estrutura diferente (a língua) e com imagens de resolução bastante diferente. Caberá notar que, nos estudos acima referenciados, foram utilizadas imagens de elevada resolução obtidas através de sequências RM com tempos de aquisição longos (acima dos 5 minutos). No nosso caso apesar de ter sido utilizada uma aquisição volumétrica, o tempo de aquisição foi de 19 s com o falante em fonação, o que condiciona, inevitavelmente, a qualidade das imagens quer no que diz respeito à S/R, quer à resolução espacial e à presença de artefactos.

A avaliação da qualidade das segmentações obtidas, a partir das técnicas de imagem, para uma determinada estrutura reveste-se de especial importância todavia, continua a constituir um enorme desafio. De acordo com Liu *et al.* (2003), quando se pretende avaliar a efetividade e utilidade de um determinado método de segmentação devem ser tidos em consideração três fatores: a precisão ou reprodutibilidade, a exatidão (*accuracy*) e a eficiência (*efficiency*). O fator mais difícil de avaliar e validar, de acordo com estes autores, é a exatidão. A exatidão avalia até que ponto a segmentação obtida por um determinado método é comparável com a verdadeira delineação do objecto (*ground truth*). Contudo, no âmbito da imagem médica, por razões óbvias, é difícil obter um *ground truth* que não seja, também ele, obtido através de técnicas de imagem. Alguns autores (Zou *et al.*, 2004), consideram que nestas situações a validação fica limitada à avaliação da reprodutibilidade da técnica utilizada. Frequentemente, mesmo com as limitações apontadas às segmentações manuais (Pham *et al.*, 2000), considera-se como referência ou *gold standard* um traçado manual efetuado por um especialista.

Também neste estudo, na impossibilidade de ter uma verdadeira referência desta estrutura (língua), se considerou como *gold standard* uma segmentação manual, efetuada por um operador com conhecimentos a nível de anatomia e familiarizado com imagens de RM, a autora

deste trabalho.

Assim, para poder efetuar a comparação entre as segmentações efetuadas com o algoritmo semi-automático foi necessário efetuar traçados manuais da língua. Estas segmentações foram conduzidas com recurso às ferramentas de segmentação manual do *ITK-SNAP*.

A segmentação manual é demorada (entre 22 a 40 minutos), como já referido na secção anterior e bastante cansativa. O processo inicia-se com a marcação de referências nos planos axial e coronal, facilitando a identificação das margens da língua, particularmente, na sua parte inferior e lateral, onde o contraste entre esta estrutura e tecidos musculares circundantes é baixo. Depois, a segmentação continua no plano sagital, corte a corte, até que toda a estrutura esteja incluída. No final, é necessário rever e corrigir a segmentação efetuada e é gerada uma malha de forma semelhante ao descrito para as segmentações efetuadas com o algoritmo de regiões de competição. A figura 5.19 apresenta um exemplo de traçado manual da língua efetuada no *ITK-SNAP*.

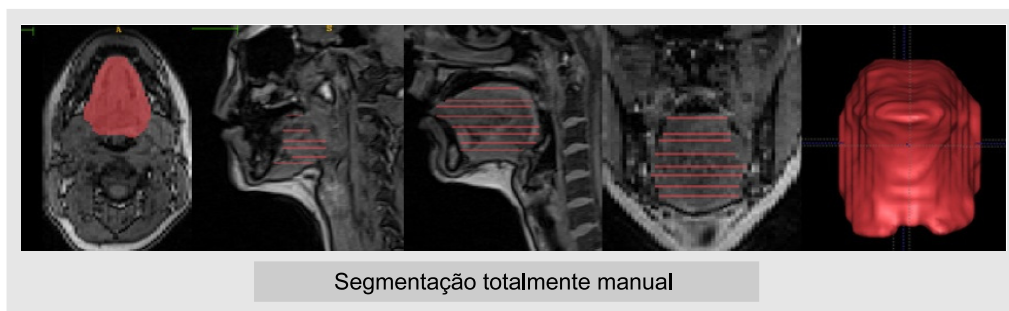


Figura 5.19: Exemplo de segmentação manual, guiada nos três planos, efetuada com o *ITK-SNAP*. Imagem publicada em Martins *et al.* (2011c).

As segmentações com utilização do algoritmo de regiões de competição implementado no *ITK-SNAP* foram também efetuadas pela autora deste trabalho. As segmentações manuais serão doravante designadas como (M) e as semi-automáticas como (SA).

Todos os procedimentos foram efetuados com um computador pessoal (Windows 7 with 6 GB RAM, 1.6 GHz Intel Core Duo processor and NVIDIA GeForce GPU). As condições ambientais foram mantidas, tanto quanto possível, constantes quando se efetuava segmentações para o processo de validação (e.g., luz ambiente). As segmentações, para este efeito, foram conduzidas ao longo de algumas semanas alternando o método de segmentação, falante e configuração da língua (som) para evitar memorização. O processo de avaliação conduzido incluiu vários tipos de análise com utilização de abordagens qualitativas (e.g., visual) e quantitativas (e.g., métricas de sobreposição). A avaliação foi considerada preliminar uma vez que contemplou apenas um observador que utilizou os dois métodos de segmentação (M *vs* SA). Desta forma, apenas se poderá avaliar a reprodutibilidade (intra-operador) de cada um dos métodos e a concordância entre estes.

5.4.1 Avaliação qualitativa

O primeiro passo neste processo, após realização de segmentações M e SA para a mesma configuração da língua, foi avaliar a qualidade geral das malhas obtidas através dos dois métodos referidos.

Nesse sentido, as malhas 3D foram apresentadas a três observadores (AS, LJ, CO)¹¹ sem que estes soubessem qual o método de segmentação utilizado para obter cada uma delas. Para cada uma de duas configurações da língua (referentes a sons diferentes) foram mostradas oito malhas (três M e cinco SA) em diferentes vistas (anterior, posterior, lateral e oblíquas). Foi solicitado que detectassem diferenças notórias entre as oito malhas, se haveria alguma que considerassem como de melhor qualidade (que se destacasse em relação às outras) e que avaliassem até que ponto conseguiram identificar a presença de *groove*, forma da língua, etc. Foram ainda apresentadas imagens, da mesma malha, com diferentes graus de suavização.

Foram identificadas, por qualquer dos observadores, diferentes formas da língua na produção de diferentes sons. Aspectos relevantes na forma e postura da língua, como a presença de depressão na linha média (*groove*), forma do dorso da língua (côncavo/convexo), posição da ponta da língua e presença ou ausência de compressão lateral em direção à linha média foram detetados nas malhas obtidas de forma M ou SA.

No que diz respeito ao grau de suavização, os observadores acharam que maior suavização conduz a uma representação mais “agradável” (note-se que as malhas com nível suavização baixo apresentam alguns artefactos em escada - *stair artifacts*) mas foram unânimes em considerar que alguns detalhes podem ser omitidos quando o grau de suavização é maior. Apenas um dos observadores (LJ) considerou como a “melhor malha” uma das resultantes de segmentação M. Esta avaliação visual, qualitativa, foi importante para perceber que não existiam diferenças, visualmente observáveis, na qualidade das malhas. Esta avaliação é, no entanto, muito subjetiva.

Um segundo passo neste processo passou pela utilização de uma ferramenta de comparação de malhas poligonais - o PolyMeco¹² (Silva, 2007; Silva *et al.*, 2009) em que se efetuou a sobreposição de cada uma das malhas (considerada a malha de referência) com todas as outras obtidas, com o mesmo método, para ter uma noção das diferenças (neste caso geométricas) entre estas.

Com recurso a uma escala de cores é possível observar se existem e onde se localizam as diferenças entre as malhas (azul escuro corresponde à ausência de diferenças e vermelho a diferenças acentuadas). Para o efeito, das várias métricas disponíveis no PolyMeco, utilizou-se a distância geométrica. Neste tipo de comparação o que efetivamente se compara são as malhas e não as segmentações propriamente ditas pelo que não é, só por si, um método suficiente para efetuar a validação da segmentação. No entanto, as pequenas diferenças observadas entre as malhas indicaram a existência de uma boa reprodutibilidade de cada um dos métodos, motivando a realização de uma análise mais aprofundada com recurso a outras métricas. Na figura 5.20 apresentamos, a título de exemplo, a sobreposição de malhas em 3D, para comparar as diferentes segmentações

¹¹A formação de dois dos observadores é na área de Engenharia Eletrónica (AT,LJ); o terceiro observador é oriundo da área da linguística (CO). Todos têm conhecimentos de fonética, estão envolvidos em estudos de produção de fala e estão familiarizados com imagens de RM obtidas neste contexto.

¹²Existem várias ferramentas para comparação de malhas poligonais que poderiam ter sido utilizadas nesta experiência como, por exemplo, o MeshDev (Roy *et al.*, 2004). A ferramenta escolhida - o PolyMeco - foi desenvolvida na Universidade de Aveiro (IEETA) por Silva (2007). É uma ferramenta de acesso livre, apresenta uma interface gráfica muito intuitiva facilitando a sua utilização por operadores que não pertencem à área da visualização ou da computação gráfica. Permite comparar várias malhas em simultâneo, lado a lado, o que é muito vantajoso. Por outro lado, ao utilizar o PolyMeco beneficiamos do apoio e sugestões do investigador que o desenvolveu, o que se revelou uma mais-valia.

SA efetuadas para cada configuração da língua. Note-se que cada imagem apresentada resulta da sobreposição de duas malhas da língua. Tendo em atenção a escala de cores, atrás referida, a diferença entre cada duas segmentações é muito subtil. Mais detalhes relativos à utilização desta ferramenta podem ser consultados no artigo apresentado no anexo E.

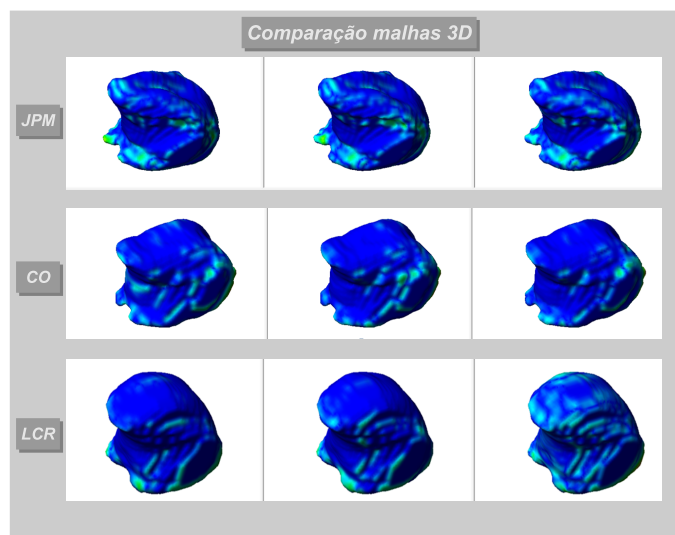


Figura 5.20: Sobreposição de malhas em 3D, para comparar as diferentes segmentações SA efetuadas. Exemplo para três falantes com produções de diferentes sons ([l] para os falantes JPM e CO e [L] para o falante LCR).

A utilização do PolyMeco, neste contexto de avaliação, conduziu-nos, posteriormente, à realização de algumas experiências (em que se compararam malhas referentes a diferentes sons, o mesmo som em contextos diferentes, entre outras comparações possíveis) com as quais se obtiveram resultados muito promissores e que de alguma forma resultaram numa aplicação, bastante inovadora, de ferramentas para comparação de malhas poligonais ao estudo dos sons da fala. Os resultados dessa aplicação, ilustrados através de alguns exemplos, encontram-se no artigo que constitui o anexo E.

5.4.2 Avaliação quantitativa

Uma avaliação qualitativa pode ser importante e fornecer uma primeira impressão acerca das segmentações efetuadas por um operador ou método de segmentação. No entanto, como referido atrás, é sempre afetada por algum grau de subjetividade introduzida por cada observador na interpretação das imagens.

Desta forma, pretendeu-se ir um pouco mais além no que diz respeito ao processo de avaliação das segmentações, recorrendo a algumas métricas (veja-se Silva *et al.*, 2011) usualmente utilizadas, no âmbito da segmentação de imagem médica, para comparação de segmentações efetuadas por algoritmos automáticos/semi-automáticos contra segmentações manuais efetuadas por operadores com experiência, comparar diferentes métodos entre si, segmentações efetuadas por diferentes observadores (inter-observador) ou, a diferentes tempos, pelo mesmo observador (intra-

observador).

5.4.2.1 Análise do volume total da língua - um falante

O passo seguinte no processo de avaliação consistiu na utilização de uma das métricas mais simples a que se pode recorrer na comparação das segmentações efetuadas por métodos diferentes ou várias vezes por um observador: o volume total da estrutura segmentada (Gerig *et al.*, 2001; Silva *et al.*, 2011).

Considerando que a língua é uma estrutura elástica, que pode assumir diferentes posturas e formas durante a produção dos diferentes sons, mas incompressível mantendo constante o seu volume durante a produção de sons (Stone e Lundberg, 1996; Takemoto, 2001; Perrier *et al.*, 2003; Badin *et al.*, 2006), poder-se-á assumir, como hipótese, que se o método de segmentação utilizado for reproduzível então os volumes obtidos, em diferentes segmentações da língua, para um mesmo falante, ao produzir diferentes sons, serão muito próximos.

Assim, ainda antes de comparar os volumes obtidos para segmentações M e SA para os vários falantes, efetuaram-se catorze segmentações da língua a partir das imagens obtidas para um dos falantes ao produzir diferentes sons ou o mesmo som produzido em contextos ou posições silábicas diferentes. Neste caso apenas foram efetuadas segmentações com o algoritmo SA. A tabela 5.5 apresenta os resultados obtidos para o volume da língua (mm^3), para as diferentes segmentações efetuadas a partir de produções da falante CO.

Tabela 5.5: Volume da língua (mm^3) obtido para 14 produções diferentes da falante CO.

Produção	Volume (mm^3)
sal	80310
sul	81575
til	77078
a	80285
i	77201
u	79903
laca	80232
litro	79832
lupa	78863
palha	77013
pilha	77632
pulha	77300
carro	81879
caro	79839
Média	79210
DP	1688

Analisando os valores contidos na tabela, verifica-se que os resultados obtidos para as catorze produções são: média=79210 mm^3 , mínimo = 77013 mm^3 , máximo = 81879 mm^3 e desvio-padrão (DP) = 1688 mm^3 . Estes resultados apontam para a conservação do volume da língua durante a produção de diferentes sons da fala indo ao encontro do reportado por Badin *et al.* (2006). Apesar de não ser das métricas mais robustas, o valor pequeno obtido para o DP (1,6 cm^3)

poderá ser interpretado como um indicador da boa reprodutibilidade do método semi-automático na segmentação das imagens da língua.

5.4.2.2 Análise do volume total da língua - vários falantes

Para avaliar a reprodutibilidade intra-operador com os dois métodos, tendo por base o volume total da língua, consideraram-se sete conjuntos de imagens (cada conjunto corresponde a uma aquisição volumétrica) relativas a diferentes configurações da língua e falantes. Cada conjunto de imagens foi segmentada duas vezes manualmente (M) e, três vezes, utilizando o algoritmo semi-automático (SA), num total de 35 segmentações da estrutura.

A tabela 5.6 resume a informação relativa aos falantes, aos sons produzidos por cada um dos falantes, ao volume total obtido para cada uma das segmentações efetuadas, aos valores médios e de DP para cada um dos métodos considerados (SA) e (M).

A análise dos resultados apresentados na tabela permite-nos constatar que o volume obtido para cada um dos falantes é diferente, o que não é de estranhar uma vez que o volume da língua se correlaciona com o biótipo do indivíduo (Ajaj *et al.*, 2005). Verifica-se também que o volume calculado mais pequeno é para a falante do género feminino. Os valores obtidos estão, no geral, de acordo com os dados reportados na literatura, no âmbito de outros estudos (Ajaj *et al.*, 2005), que incluem dados referentes ao volume da língua para indivíduos do género feminino e masculino.

De salientar que analisando as duas tabelas, no que diz respeito à falante CO, verifica-se que os valores de DP para as diferentes segmentações relativas à mesma configuração da língua (ver tabela 5.6) são menores que os obtidos quando se estimou o volume para as catorze configurações relativas a diferentes sons (tabela 5.5). Muito provavelmente a variação da forma introduz mais variabilidade que a introduzida pelo observador ou pelo algoritmo nas diferentes segmentações para a mesma configuração (mesmo som). Os volumes obtidos a partir dos traçados manuais, para o mesmo som e falante, são similares aos obtidos com o método SA com uma tendência para serem ligeiramente inferiores.

Tabela 5.6: *Volume da língua (mm^3) obtido para vários falantes. Valor médio e desvio padrão (DP) para os métodos manual (M) e semi-automático (SA).*

<i>Falante</i>	<i>Som</i>	<i>SA1</i>	<i>SA2</i>	<i>SA3</i>	<i>Média</i>	<i>DP</i>	<i>M1</i>	<i>M2</i>	<i>Média</i>	<i>DP</i>
CO	sal	79836	80245	80110	80064	208	76816	77418	77117	426
CO	sul	81567	82862	79947	81459	1461	80026	77988	79007	1441
CO	vogal a	80143	80003	80161	80102	86	80428	79550	79989	621
JPM	laca	140966	143417	142149	142177	1226	137366	139271	138319	1347
JPM	vogal a	143346	143077	141622	142682	928	132852	133775	133314	653
LCR	laca	130006	129372	124700	128026	2898	121866	119931	120899	1366
LCR	palha	126972	125490	126583	126348	768	121025	120950	120988	53

De forma análoga ao efetuado em estudos de validação com o ITK-SNAP (Yushkevich *et al.*, 2006; Ho *et al.*, 2003), foi calculado o Coeficiente de Correlação Intra-classes (CCI). Os resultados obtidos quando comparadas as diferentes classes apresentam-se na tabela 5.7. Os valo-

res, igualmente altos, obtidos para os pares M×M e SA×SA apontam para uma reprodutibilidade elevada e equivalente, para ambos os métodos.

Tabela 5.7: Resultados obtidos após cálculo do coeficiente de correlação intra-classes (CCI).

Pares	CCI
M×M	0.99
SA×SA	0.99
M×SA	0.98

5.4.2.3 Métricas de semelhança (*Jaccard e Voxelwise difference*) - informação espacial

Apesar da análise baseada no volume total poder dar uma boa indicação relativa à reprodutibilidade do método de segmentação, na realidade, pode acontecer que duas segmentações apresentem exatamente o mesmo volume mas com uma sobreposição espacial fraca. No limite pode nem existir sobreposição das segmentações (Gerig *et al.*, 2001).

Para avaliar de forma mais fiável quão bem duas segmentações são concordantes é necessário recorrer a métricas mais robustas que incorporem também informação espacial - designadas *overlapping metrics* ou métricas de sobreposição. Nos estudos consultados uma das métricas mais referidas e utilizadas na prática é o coeficiente de semelhança de *Dice* (Zou *et al.*, 2004; Yushkevich *et al.*, 2006) mas muitas outras estão disponíveis (veja-se, por exemplo Gerig *et al.*, 2001 ou Silva *et al.*, 2011). A escolha das métricas a utilizar é relevante e nem sempre é simples pois, muitas vezes, a informação fornecida pode não ser a adequada ou, entre as várias métricas, a informação pode ser redundante.

Nesse sentido, tendo por base um estudo preparatório efetuado por Silva *et al.* (2011), realizado noutro contexto (validação de segmentações de imagem cardíaca), foram escolhidas para esta avaliação duas métricas - o coeficiente de *Jaccard* e *Voxelwise difference* - que, segundo os autores, nos podem fornecer informação complementar.

A figura 5.21 apresenta um esquema que ilustra as regiões de interesse quando se pretende analisar o grau de semelhança e sobreposição entre as duas regiões X e Y. As formulações matemáticas referentes às métricas utilizadas estão também incluídas na imagem.

O coeficiente de *Jaccard* é igual a 1 se existir uma sobreposição perfeita das duas regiões e 0 se não existir sobreposição. A outra métrica (*voxelwise difference*) calcula a razão entre o n^o de falsos positivos (FP) e falsos negativos (FN) e o volume médio segmentado (considerando as segmentações que estão a ser comparadas).

Para esta análise foram incluídos todos os pares (M *vs* M, SA *vs* M e SA *vs* SA), num total de 70 comparações. A tabela 5.8 apresenta os resultados obtidos (mínimo, máximo, média e desvio padrão) para ambas as métricas, para todos os pares de comparações. Na figura 5.22 apresentam-se os *boxplots* correspondentes.

Foi efetuada a análise de variância (*One-way ANOVA*) considerando, como fator, o mé-

X é a região comparada com Y (referência)

True positives (TP)	a	$ X \cap Y $
False positives (FP)	$b1 + b2$	$ X - X \cap Y $
False negatives (FN)	$c1 + c2$	$ Y - X \cap Y $
True negatives (TN)	d	$ X \cup Y $

$$Jac = \frac{|X \cap Y|}{|X \cup Y|} \quad V = \frac{FP + FN}{MeanVolume} = \frac{|X| - |X \cap Y| + |Y| - |X \cap Y|}{(|X| + |Y|)/2}$$

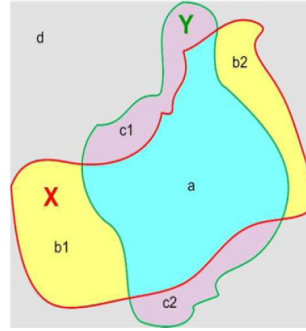


Figura 5.21: Áreas notáveis consideradas na análise de semelhança entre as regiões X e Y . As fórmulas de cálculo para o coeficiente de *Jaccard* e *Voxelwise difference* são apresentadas na parte inferior da figura. Imagem adaptada de Silva *et al.* (2011) com permissão dos autores.

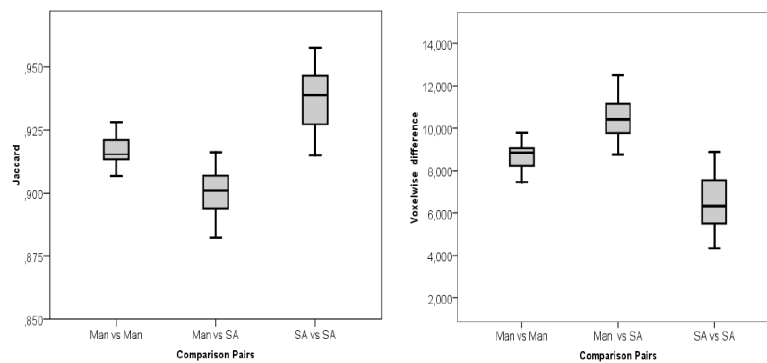


Figura 5.22: *Boxplots* para as métricas de sobreposição, para todos os pares comparados. À esquerda, coeficiente de *Jaccard* [0,1]; *voxelwise difference* (%), à direita.

Tabela 5.8: *Mínimo, Máximo, Média e DP obtidos para as duas métricas a partir de todos os pares comparados. O coeficiente de Jaccard varia entre [0, 1] e o Voxelwise difference é apresentado em (%).*

Métrica	Pares	Min	Máx	Média	DP
<i>Jaccard</i>	MvsM	0.9066	0.9281	0.9170	0.0075
<i>Voxelwi</i>	MvsM	7.4586	9.7986	8.6624	0.8101
<i>Jaccard</i>	MvsSA	0.8822	0.9162	0.9002	0.0079
<i>Voxelwi</i>	MvsSA	8.7477	12.5098	10.4950	0.8730
<i>Jaccard</i>	SAvsSA	0.9150	0.9576	0.9382	0.0123
<i>Voxelwi</i>	SAvsSA	4.3371	8.8769	6.3876	1.3111

todo de segmentação (SA, M). Antes da aplicação do teste foram verificados os pressupostos de normalidade (teste de *Kolmogorov-Smirnov*) e de homogeneidade (teste de *Levene*). A tabela 5.9 resume os resultados obtidos. O tratamento estatístico foi efetuado com o SPSS 15.0 (SPSS Inc., Chicago, IL, USA).

Tabela 5.9: *Resultados ANOVA relativos às métricas (Jaccard e Voxelwise difference), considerando o método de segmentação como fator.*

Métrica	Mean Sq	<i>F</i>	<i>p-value</i>
<i>Jaccard</i>	0.002	18.232	.001
<i>Voxelwi</i>	27.166	18.433	.001

5.4.2.4 Discussão dos resultados relativos ao estudo de avaliação

O método de segmentação semi-automático por nós utilizado para segmentação da língua, com recurso ao ITK-SNAP, tendo por base o algoritmo que recorre às regiões de competição (*Region Competition Snakes*) e que foi objeto desta avaliação preliminar mostrou ser muito mais eficiente que o método de segmentação, totalmente manual, efetuado por um operador experiente. O tempo de segmentação é consideravelmente mais baixo, sendo também, muito menor, o esforço e cansaço do operador na tarefa de segmentação. Para análise de bases de dados de grandes dimensões esta abordagem constitui com toda a certeza uma mais-valia permitindo obter resultados num tempo razoável. No entanto, torna-se necessário avaliar a reprodutibilidade do método e a sua precisão.

A avaliação quantitativa baseada na comparação do volume total apontou para uma excelente reprodutibilidade dos dois métodos. No entanto, esta métrica não tem em consideração diferenças espaciais. Os resultados da análise de variância mostraram que existe uma diferença estatisticamente significativa entre os dois métodos de segmentação (*Jaccard* $F=18.232$, $p<.001$; *voxelwise* $F=18.433$, $p<.001$) - esta diferença deve-se ao facto de o método semi-automático (*Jaccard*=0.9382 e 6.3876 *voxel error*) ser ainda mais reprodutível que o método manual (*Jaccard*=0.9170 e 8.6624 *voxel error*).

Os valores de semelhança obtidos através do coeficiente de *Jaccard* e *Voxelwise difference* apontaram também para bons níveis de concordância entre as abordagens semi-automática e manual (*Jaccard*=0.9002 and 10.4950 *voxel error*).

O valor ligeiramente mais baixo para o coeficiente de *Jaccard* observado quando se compararam os pares (M vs SA) deve-se muito provavelmente a uma segmentação mais conservadora nos traçados manuais o que está de acordo com o que tinha sido obtido considerando o volume total como métrica.

Os resultados, no que respeita à reprodutibilidade intra-observador para as segmentações da língua, estão em linha com os resultados de validação já reportados, para o *ITK-SNAP*, na segmentação de outras estruturas anatómicas ou patologias (e.g., núcleo caudado, ventrículos laterais ou tumores cerebrais). No global, o desempenho do algoritmo que recorre às regiões de competição surpreendeu-nos de forma muito positiva uma vez que a qualidade dos volumes adquiridos, dado o tempo de aquisição curto, é consideravelmente inferior ao dos estudos que nos serviram de referência. Assim, dada a sua maior eficiência e boa reprodutibilidade, o método SA pode constituir uma alternativa viável à segmentação manual da língua.

Apesar dos resultados serem encorajadores, limitam-se a uma avaliação da reprodutibilidade intra-observador. Será necessário continuar esta avaliação incluindo pelo menos mais um observador o que permitirá avaliar a variabilidade entre observadores, aumentar o número de pares de segmentações, considerar outros fatores (e.g., ordem da segmentação, falante) e interação entre estes para um tratamento estatístico mais robusto. Se, como esperamos, tendo em consideração o reportado nos estudos referidos, os resultados forem igualmente bons com mais observadores, este método de segmentação implementado no *ITK-SNAP* pode vir a ser útil, em contexto clínico, para obtenção de volumes da língua, por exemplo, em pacientes com patologias como apneia do sono, macroglossia, etc.

5.5 Resultados

Nas secções seguintes, serão apresentados os resultados mais relevantes extraídos do conjunto de dados de RM adquiridos e que contemplam as consoantes laterais. A caracterização articulatória das laterais será efetuada, tendo como base todo o tipo de informação disponibilizada pela aplicação das metodologias anteriormente descritas. Assim, contaremos com informação de natureza qualitativa, como a proveniente da observação direta das imagens (nos vários planos), da sobreposição e comparação de contornos 2D, de visualizações e sobreposição de malhas em 3D, etc. Esta informação será consubstanciada com dados quantitativos resultantes do cálculo dos IP sobre os contornos 2D, informação relativa às áreas, funções de área do trato vocal e das passagens laterais. Os resultados serão apresentados de forma separada, primeiro para a lateral alveolar /l/ e, seguidamente, para a palatal /L/. No que diz respeito à palatal, serão também incluídos os resultados de uma primeira experiência de simulação efetuada. No final, serão apresentadas algumas medidas referentes ao comprimento do trato vocal.

5.5.1 A lateral alveolar - /l/

A descrição articulatória do /l/ incluirá a referência ao ponto e extensão da articulação, à forma da língua em 2D e 3D, às áreas ao longo do trato vocal e à dimensão das passagens laterais. Posteriormente, serão apresentados os dados que nos permitirão aferir a influência da posição silábica e do contexto vocálico na produção da consoante.

5.5.1.1 Ponto e extensão do contacto alveolar na linha média

Uma vez que a divisão do palato e da língua em zonas nem sempre é consensual e é bastante variável nas referências consultadas, optou-se por seguir, nesta descrição, a divisão e terminologia utilizada por Recasens *et al.* (1993) e Recasens *et al.* (1996) e que se encontra representada na figura 5.23.

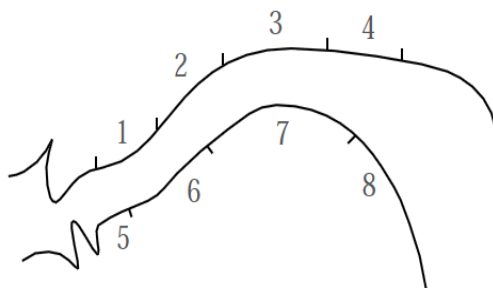


Figura 5.23: Representação do trato vocal com as zonas e sub-zonas articulatórias e regiões da língua: (1) zona alveolar, (2) pré-palatal, (3) médio-palatal, (4) pós-palatal, (5) ápice e lâmina lingual, (6) pré-dorso, (7) parte média do dorso e (8) pós-dorso.

No que diz respeito ao ponto de articulação da lateral alveolar, tendo por base a observação de imagens 2D como as que se apresentam na figura 5.24, verifica-se, sem grandes surpresas, que esta é produzida com contacto da lâmina lingual (articulação laminal) ou com o ápice da língua (articulação apical) na linha média, na região dento-alveolar.

Na verdade, a comparação com as zonas articulatórias consideradas na figura 5.23 sugere que as articulações produzidas por estes falantes são bastante anteriores, algumas não ocorrendo plenamente na zona alveolar mas, mais adiante, na zona dental¹³.

As articulações laminais foram, para este grupo de falantes, mais frequentes (70.0%) do que as articulações apicais (28.5%). Numa produção não houve contacto com a região alveolar. Os falantes JH e MC foram os que apresentaram, com mais frequência, articulações apicais (66.6%). Do conjunto de todas as produções, não se verificou qualquer tendência ou padrão que indique uma associação entre o tipo de articulação e a posição silábica.

¹³Nas imagens apresentadas não estão co-registadas as arcadas dentárias. Contudo, é possível verificar que, em muitas produções a ponta da língua entra em contacto com uma zona de vazio de sinal correspondente aos incisivos superiores (cf. figura 5.24, para as produções da falante MC e do falante JH).

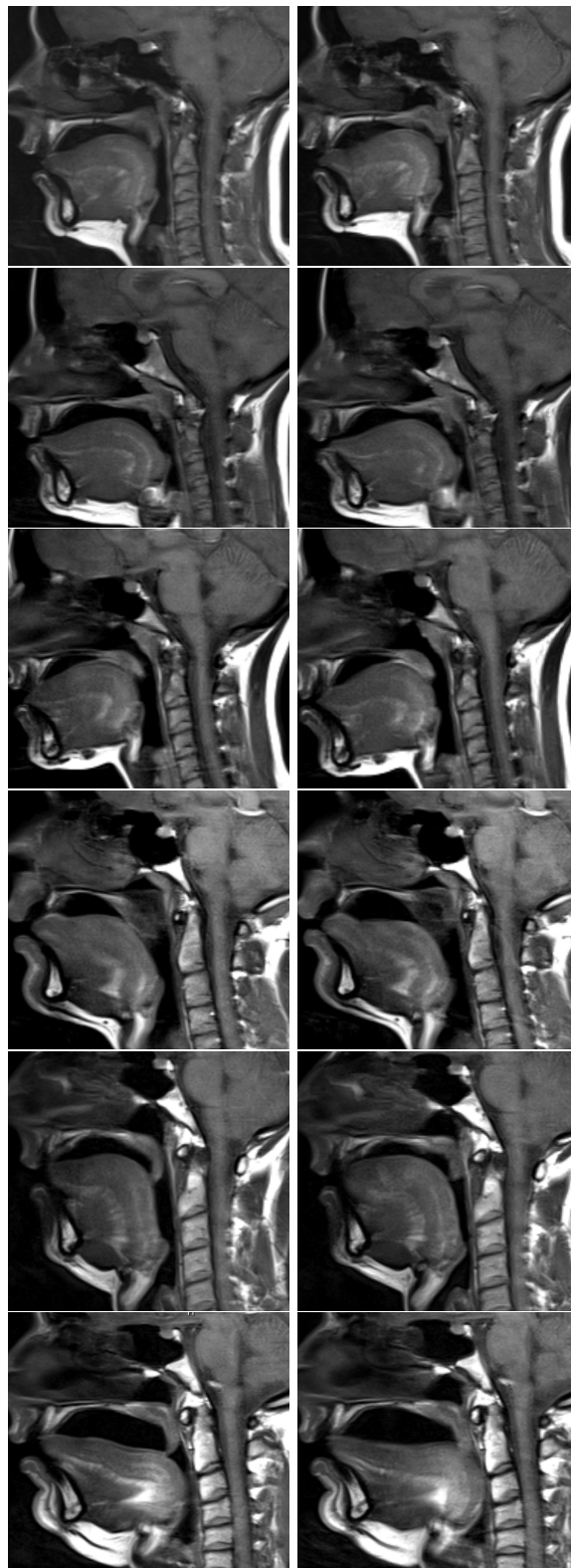


Figura 5.24: Imagens 2D no plano médio-sagital relativas a seis falantes durante a produção do /l/ em posição de ataque (esquerda) e em coda (direita). Ponto de articulação da lateral e configuração da língua e trato vocal em 2D. De cima para baixo: falantes CO, MC, ER, LCR, JH e JPM.

Para além do local de contacto, é importante para a caracterização da articulação da lateral, obter informação relativa à extensão desse contacto na linha média. A tabela 5.10 apresenta uma estimativa da extensão dos contactos na linha média para o /l/ em ataque (como, por exemplo, em [lak6]), ataque intervocálico (como, por exemplo, em [sal6]) e coda (como em [sal]), no contexto das três vogais. A tabela inclui dados referentes a todos os sete falantes participantes. A informação proveio das produções adquiridas em 3D e foi extraída de forma automática a partir das segmentações efetuadas para obter as áreas ao longo do trato vocal¹⁴.

Do conjunto de todas as produções efetuadas pelos falantes para o /l/, verificou-se ausência de contacto alveolar em uma das produções do falante JH ocorrida em posição de coda ([l] como em [sal]) o que representa cerca de 1.5% das produções consideradas ($n = 63$). A falante MC apresentou, para o mesmo contexto e posição silábica, uma extensão do contacto muito pequena (2.0 mm).

Para as restantes produções dos falantes, a extensão do contacto na linha média variou entre os 3.0 e os 13.7 mm (maior valor para o falante JPM em [lit4u]). Das falantes do género feminino, a falante MC foi a que apresentou os contactos mais curtos, enquanto que dos masculinos foi o falante JH. Será de realçar que foram precisamente estes dois falantes que mais vezes produziram articulações apicais, como se referiu no parágrafo anterior. O contacto alveolar foi, geralmente, mas nem sempre, ligeiramente mais curto no /l/ produzido em posição de coda do que quando produzido em posição de ataque.

5.5.1.2 Configuração da língua em 2D e 3D

Para além do ponto de articulação, é importante descrever a configuração e postura assumida pela língua na produção da consoante. As imagens apresentadas na figura 5.24, que se referem a uma única imagem (2D) adquirida no plano médio sagital, permitem-nos observar a forma da língua em 2D. Como se pode constatar, existe alguma variabilidade, entre os falantes, na postura assumida pela língua na produção do /l/. Por exemplo, nas falantes CO, MC e no falante AS o ápice da língua está mais elevado (em ataque e em coda) do que a parte média do dorso da língua e forma-se depois uma elevação da região posterior do dorso. No falante LCR, também nas duas posições, o ápice da língua está mais elevado, existe uma pequena depressão logo atrás da zona de contacto após o que a posição do dorso lingual baixa gradualmente. Já no falante JH o ápice e o dorso da língua estão aproximadamente ao mesmo nível. No falante JPM, verifica-se, apenas na posição de coda, uma depressão após o zona de contacto e mais posteriormente uma ligeira elevação do dorso em direção à zona velar. Para as outras produções efetuadas pelos falantes o padrão observado é semelhante.

Com frequência dois dos falantes (ER e AS) efetuaram as produções do [l] com o véu

¹⁴Numa primeira fase deste trabalho, a extensão do contacto alveolar foi obtida através da análise direta das imagens, contando o número de cortes onde se visualizava contacto da língua com a zona alveolar. O valor do comprimento era depois calculado multiplicando o número de cortes pela dimensão do *voxel* nessa direção (dimensão do *voxel* oscila entre 0.994 e 0.999). Os resultados obtidos através dos dois métodos revelaram-se similares, sendo que este método é muito mais laborioso. A extração automática dos dados, por outro lado, é significativamente mais rápida, mas nalgumas situações apresentou falhas. Por essa razão, as medidas incluídas na tabela relativas aos falantes ER e AS foram obtidas de forma manual.

Tabela 5.10: *Extensão dos contactos na linha média para todos os falantes. Valores para o /l/ em ataque, ataque intervocálico (At.inter) e coda.*

Falante	Vogal	/l/ Ataque (mm)	/l/ At. inter (mm)	/l/ Coda (mm)
CO (F)	/a/	9.6	5.2	4.7
	/i/	5.0	8.8	9.0
	/u/	7.3	3.6	6.1
MC (F)	/a/	3.3	3.0	2.0
	/i/	5.5	3.1	3.6
	/u/	5.3	5.4	5.9
ER (F)	/a/	8.9	7.0	7.9
	/i/	7.9	6.0	6.9
	/u/	7.9	6.0	9.0
JH (M)	/a/	9.6	8.7	0.0
	/i/	5.4	4.8	5.7
	/u/	7.7	4.8	5.7
LCR (M)	/a/	10.9	6.5	10.1
	/i/	9.3	9.6	8.7
	/u/	12.5	6.1	9.1
JPM (M)	/a/	7.7	9.3	9.1
	/i/	13.7	3.1	9.5
	/u/	9.4	8.3	5.3
AS (M)	/a/	7.0	5.0	9.0
	/i/	7.9	8.9	7.9
	/u/	5.9	9.9	6.9

palatino numa posição baixa (o falante AS não está representado na figura). Em duas produções, uma do falante JH (em [sul]) e outra do falante AS, houve formação de uma pequena cavidade sub-lingual.

A configuração da língua em 3D é importante para a caracterização das sons em geral e torna-se particularmente relevante para o estudo das consoantes laterais (Stone e Lundberg, 1996; Narayanan *et al.*, 1997a, 1999; Badin *et al.*, 2002; Birkholz, 2013). A informação obtida através dos volumes de dados adquiridos e seu processamento permite-nos obter informação diversificada e bastante detalhada no que diz respeito a esta estrutura. Desta forma procurou-se efetuar uma descrição da configuração da língua em 3D, considerando a forma do seu dorso (côncava ou convexa), a existência ou não de *groove* na linha média e a existência ou não de compressão medial ou estreitamento da língua.

Esta caracterização 3D da língua provém essencialmente da observação direta das imagens no plano coronal, de visualizações 3D da língua e de malhas superficiais criadas a partir das segmentações efetuadas, para esta estrutura, com recurso ao *MeVisLab* e ao *ITK-SNAP*. Quando analisada a informação 3D, verifica-se que, apesar das diferenças entre falantes, no que diz respeito à forma da língua em 2D, há alguns aspetos que são comuns a todos falantes e para as diferentes produções efetuadas.

A figura 5.25 apresenta imagens coronais relativas a dois falantes masculinos (JPM e JH) e dois femininos (ER e CO). À esquerda, a imagem refere-se à lateral em posição inicial no

contexto da vogal [a] e, à direita, durante a produção da vogal [a]. A comparação da consoante lateral com outro som (vogal) facilita a visualização de algumas características da língua a que nos referiremos.

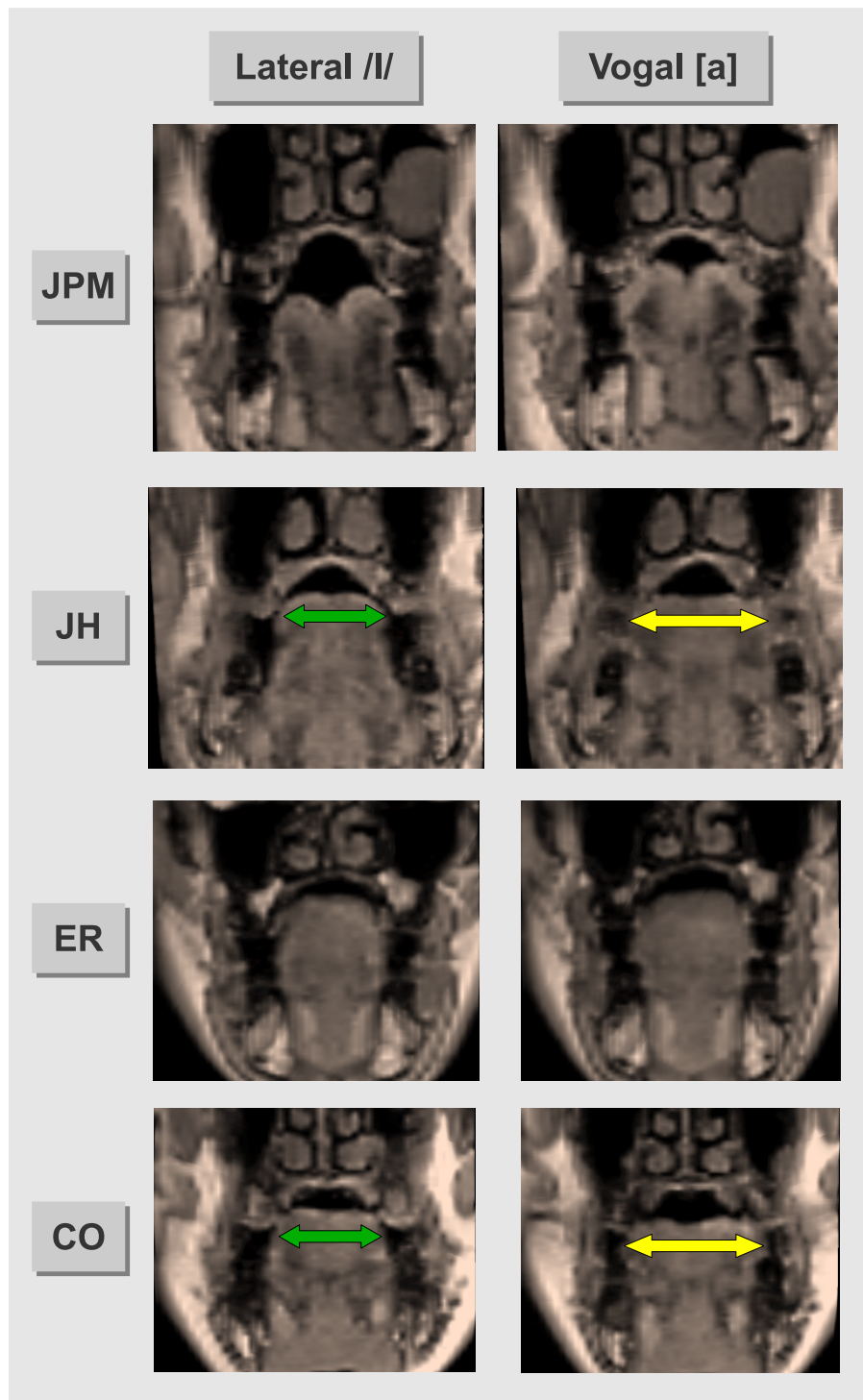


Figura 5.25: Imagens da língua no plano coronal durante a produção da lateral em posição de ataque [lak6] no contexto de [a] (à esquerda) e na produção da vogal [a] (à direita).

Como se pode observar na figura, é evidente a existência de algum grau de compressão da língua em direção à linha média. A dimensão e extensão desta compressão é que é variável entre os falantes com repercussões na dimensão (área e extensão) das passagens laterais que se estabelecem de um ou dos dois lados da língua na produção das consoantes laterais (ver secção 5.5.1.4). No caso das falantes do género feminino, a compressão da língua é, tendencialmente, mais anterior (na zona da lâmina), enquanto que nos falantes masculinos se estende mais para trás abrangendo o dorso lingual. Note-se também que o dorso da língua apresenta forma convexa com os bordos laterais da língua ligeiramente abaixados.

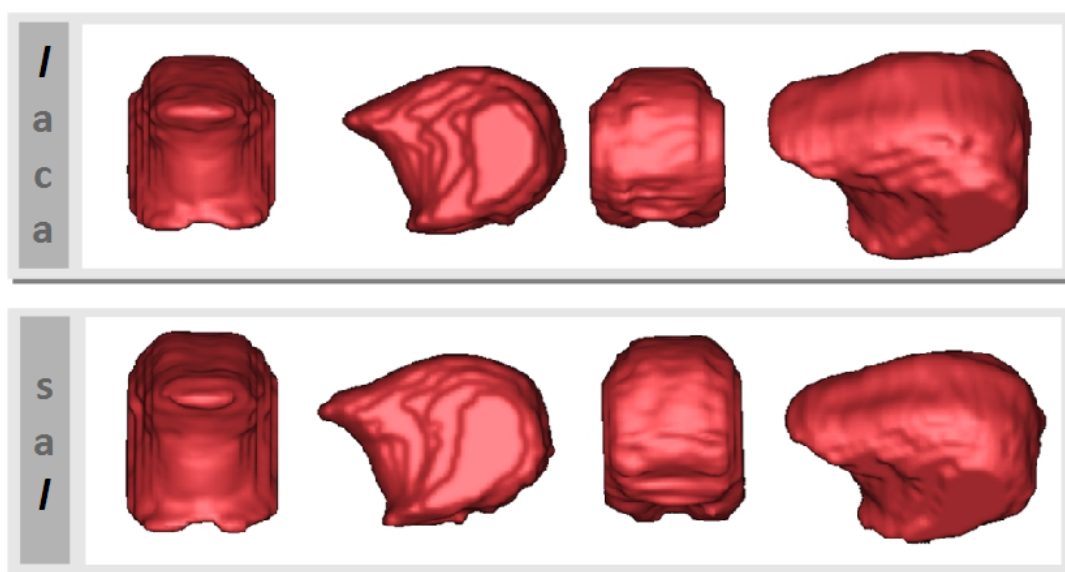


Figura 5.26: Malhas 3D da língua obtidas com o *ITK-SNAP*. Em cima, /l/ produzido, pela falante CO, em ataque e, em baixo, em posição de coda.

A figura 5.26 apresenta malhas 3D da língua, em várias vistas, obtidas para a falante CO, para as produções do /l/ em [lak6] (em cima) e [sal] (em baixo). Nesta figura, é também possível observar a forma convexa do dorso da língua. Esta característica manifesta-se, de forma semelhante, para as duas posições silábicas.

Um outro aspeto que é habitualmente considerado quando se descreve a forma do dorso da língua é a formação, ou não, de *groove* a nível da linha média. No caso de sons como, por exemplo, algumas consoantes fricativas a presença de *groove* é uma condição essencial à sua produção (Narayanan *et al.*, 1996; Bone *et al.*, 2011).

A figura 5.27 ilustra, de forma bastante clara, a presença de *groove* ao longo da parte média e pós-dorsal da língua, numa das produções ([lak6]) do falante masculino JPM. No entanto, esta característica, frequente para este falante, não foi observada de forma sistemática para os restantes falantes, pelo menos de forma tão acentuada. Em duas das produções dos falantes JH e LCR há também formação de um pequeno *groove* na região posterior ao contacto alveolar mas muito mais subtil. Note-se que na figura 5.26, para a falante CO, não se observa a formação de *groove*, apresentando o dorso uma forma mais plana (*flattening* é o termo usualmente utilizado na

literatura em língua inglesa em confronto com *grooving*).

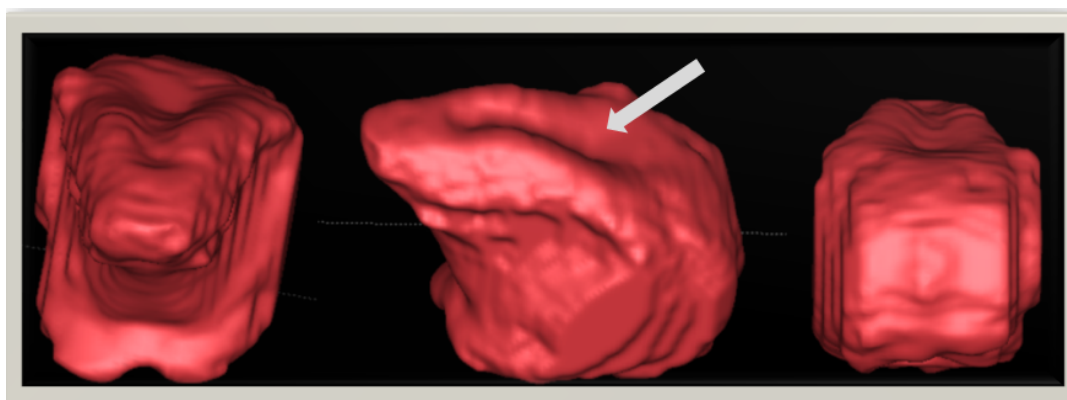


Figura 5.27: Malha superficial da língua onde se observa a presença de *groove* ao longo da linha média. Imagem referente ao falante JPM na produção do /l/, em ataque.

5.5.1.3 Funções de área do trato vocal

O espaço de propagação do ar ao longo do trato vocal define a chamada função de área do trato vocal e está diretamente relacionada com a configuração da língua em 3D. A função de área fornece o padrão de variação da área ao longo do trato vocal. O processo de extração das funções de área do trato vocal é sempre, independentemente do método utilizado, um processo empírico e bastante complexo resultando em aproximações do objeto real (Narayanan *et al.*, 1997a; Zhou, 2009). As áreas podem ser estimadas a partir das distâncias médio-sagitais quando não existe informação 3D (por exemplo, a partir de dados de cine-radiografia ou RM-2D) ou através de aplicação de uma grelha adaptativa fixa ao longo do trato vocal.

Neste estudo, as áreas do trato foram obtidas a partir de cortes reconstruídos (reconstrução multiplanar curva). Estes cortes serão aproximadamente, em cada ponto, perpendiculares ao ponto médio (*centerline*) do trato vocal e, desta forma, as áreas obtidas constituirão uma aproximação mais fidedigna do cenário real. Veja-se a secção 5.3.4.1 das metodologias onde se descrevem todas as etapas que conduziram à obtenção destes dados.

As figuras 5.28 e 5.29 apresentam alguns exemplos dos cortes a partir dos quais se obtiveram as funções de área do trato vocal, referentes à lateral em posição de ataque como em [lak6], para os falantes JPM e CO. A figura 5.30 apresenta informação similar para o falante LCR, para o [l] em posição de ataque, mas agora no contexto da vogal [i] como em [lit4u].

Como referimos, as funções de área do trato vocal foram obtidas a partir dos cortes segmentados. Tendo em atenção a dimensão da base de dados adquirida, não seria viável apresentar todas as figuras e imagens que nos permitiram efetuar a caracterização destes sons. Assim, as imagens que se mostram foram selecionadas com o intuito de facilitar a leitura e ilustrar alguns aspetos que queremos destacar ao longo do texto. No que diz respeito às funções de área, apresentaremos informação resultante da análise das produções de seis falantes, três do género feminino e três do género masculino.

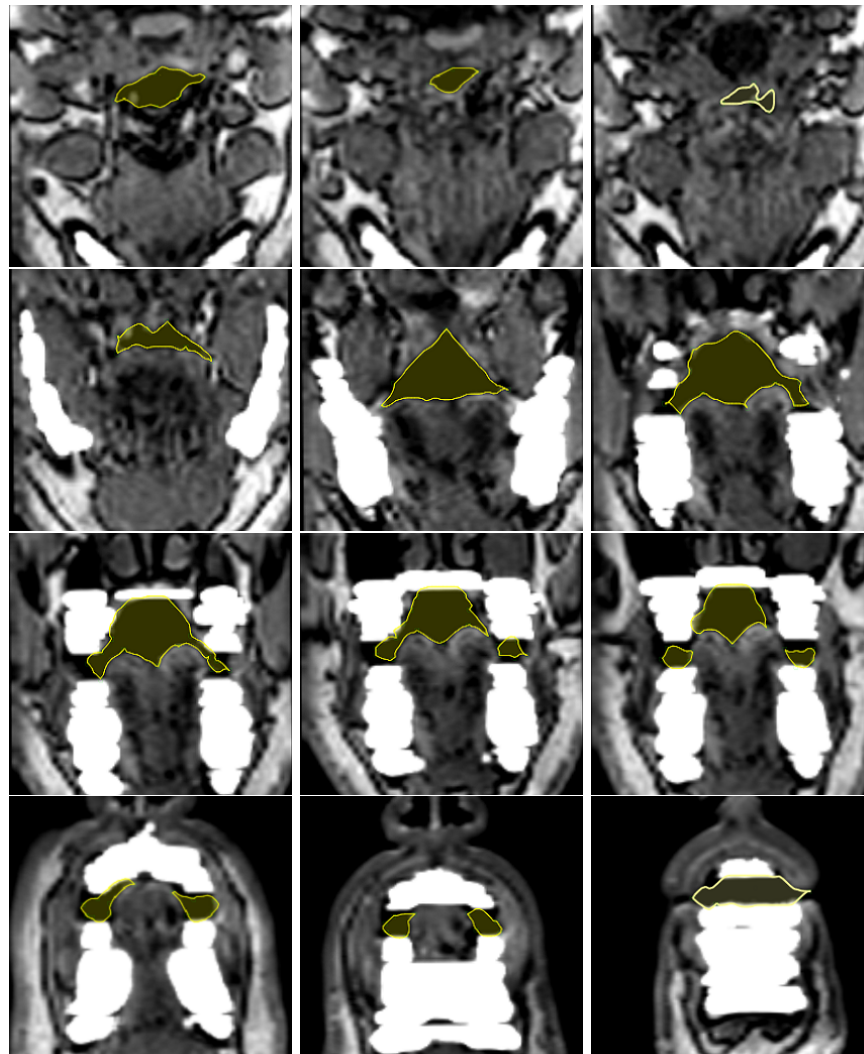


Figura 5.28: Áreas segmentadas ao longo do trato vocal para o falante JPM, na produção de [lak6].

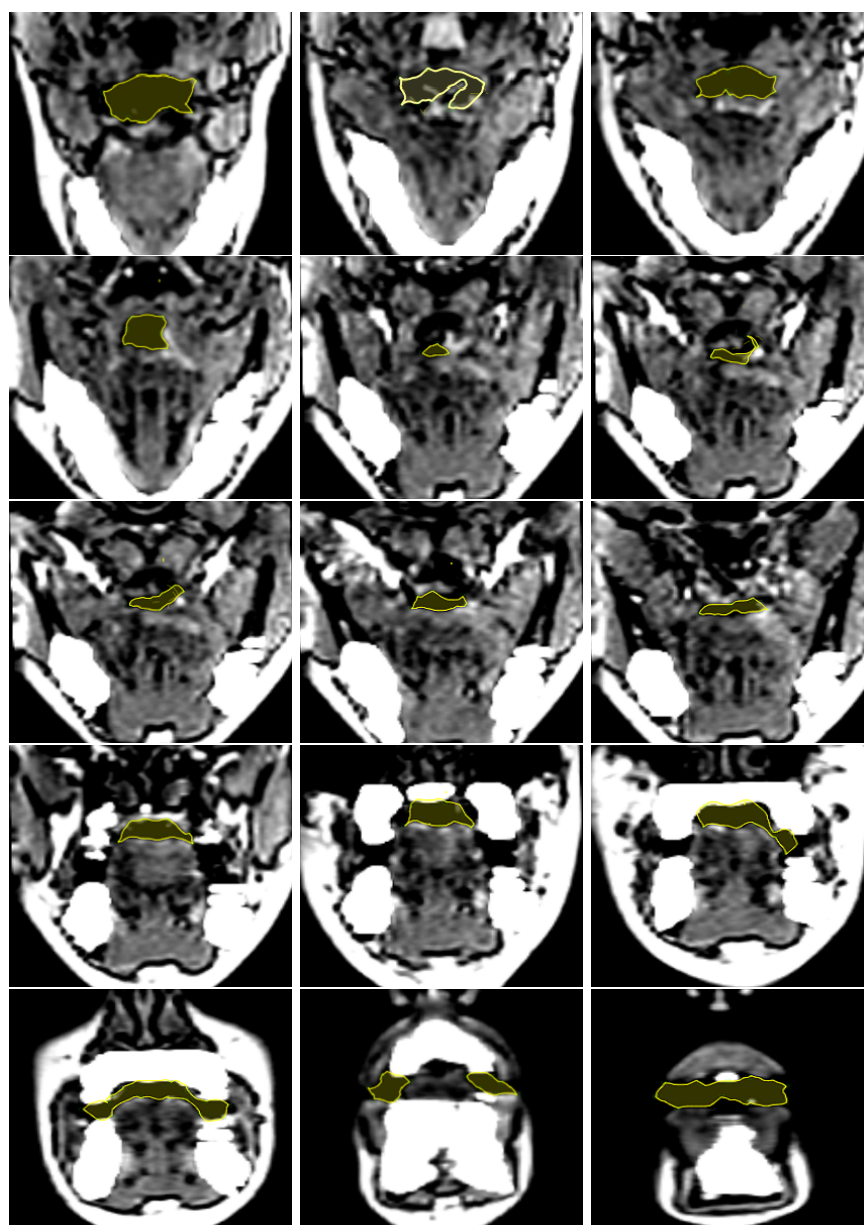


Figura 5.29: Áreas segmentadas ao longo do trato vocal para a falante CO, na produção de [lak6]

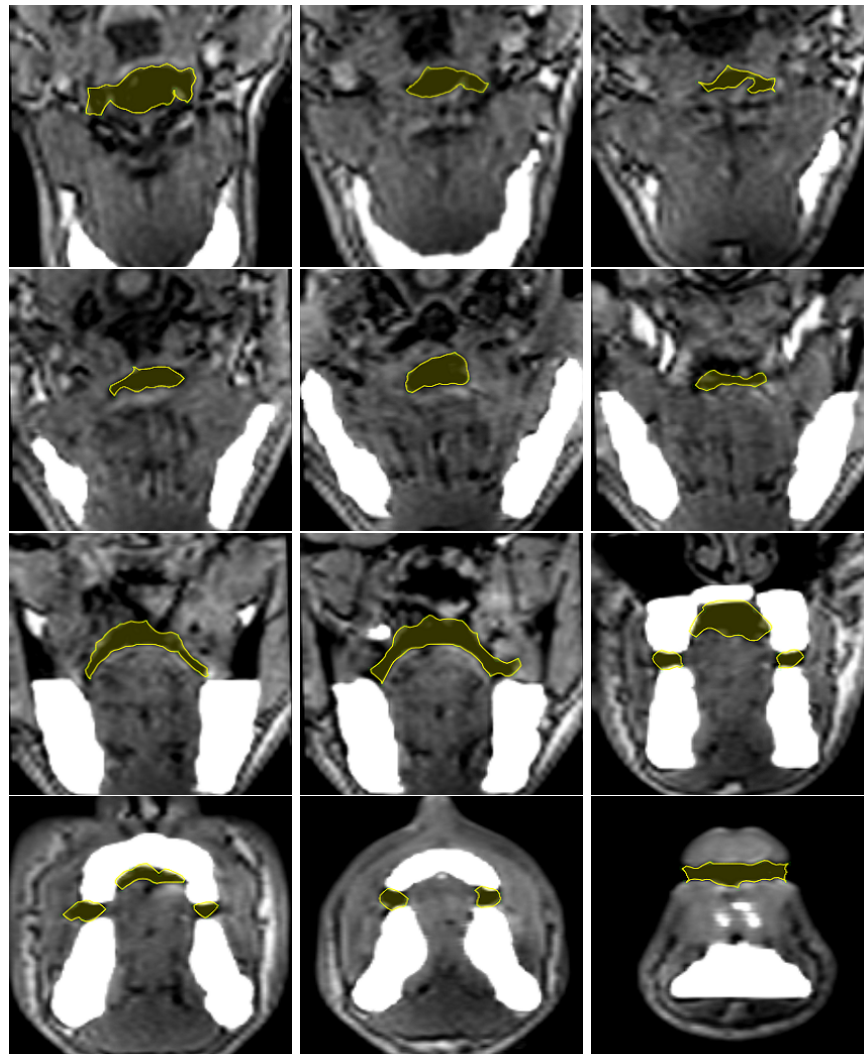


Figura 5.30: Áreas segmentadas ao longo do trato vocal para o falante LCR, na produção de [lit4u].

A figura 5.31 apresenta um exemplo de função de área referente ao falante JPM na produção da lateral em ataque intervocálico no contexto de [i] como em [biliS]. Na figura, estão identificadas algumas zonas do trato vocal para facilitar a sua descrição. A informação contempla as áreas (em cm^2) desde a zona da abertura labial (à esquerda) até à glote (à direita).

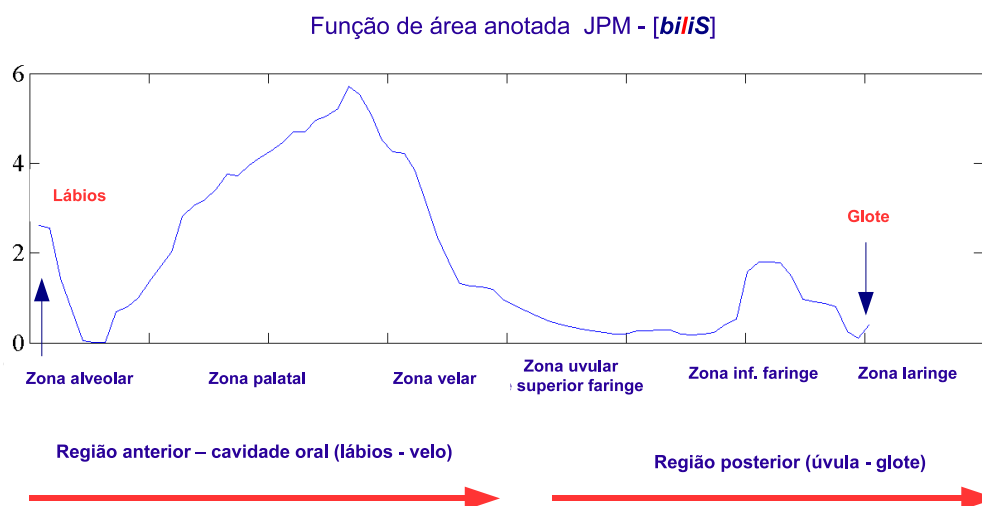


Figura 5.31: Função de área anotada referente ao /l/ para o falante JPM, no contexto da vogal [i], em posição de ataque intervocálico [biliS].

Como se pode observar, percorrendo o trato vocal desde os lábios até à glote, verifica-se que um pouco atrás da abertura labial (cerca de 1 cm) há uma zona onde as áreas são nulas que corresponde à zona do contacto alveolar, na linha média. Na zona posterior ao contacto, as áreas vão aumentando, de forma gradual, na zona palatal, atingindo um valor máximo (para este falante) de aproximadamente 6 cm^2 ; depois as áreas decrescem ao longo da região velar sendo muito baixas (0.5 cm^2) ao longo da faringe. Na zona inferior da faringe e região supra-glótica, as áreas aumentam um pouco (até 2 cm^2) diminuindo depois até à glote. As áreas de valores mais elevados na zona média e posterior do palato devem-se, muito provavelmente, ao contributo do fluxo de ar que passa lateralmente pelos bordos da língua (uma vez que esta está comprimida e com os bordos laterais baixos), mas sem que toque lateralmente no palato (veja-se a imagem à esquerda na linha 3 da figura 5.28 noutra produção do mesmo falante).

As imagens da figura 5.32 apresentam as funções de área para quatro falantes para outra produção do /l/ ([lak6]). Uma análise global permite-nos constatar que o padrão de variação de áreas é semelhante com áreas maiores a nível da cavidade oral (ao longo do palato) e menores na zona velar (falantes femininas) ou ao longo da faringe (falantes masculinos). A magnitude das áreas é diferente, entre os falantes, uma vez que a dimensão do trato vocal varia para cada um deles.

A análise e comparação das funções de área de cada um dos falantes torna-se demorada e é complicada de analisar devido às naturais diferenças de área e comprimento do trato vocal. As imagens com sobreposição das funções de área de todos os falantes para uma determinada produção revelou-se difícil de interpretar. Para ultrapassar este problema, e de modo a conseguir obter uma

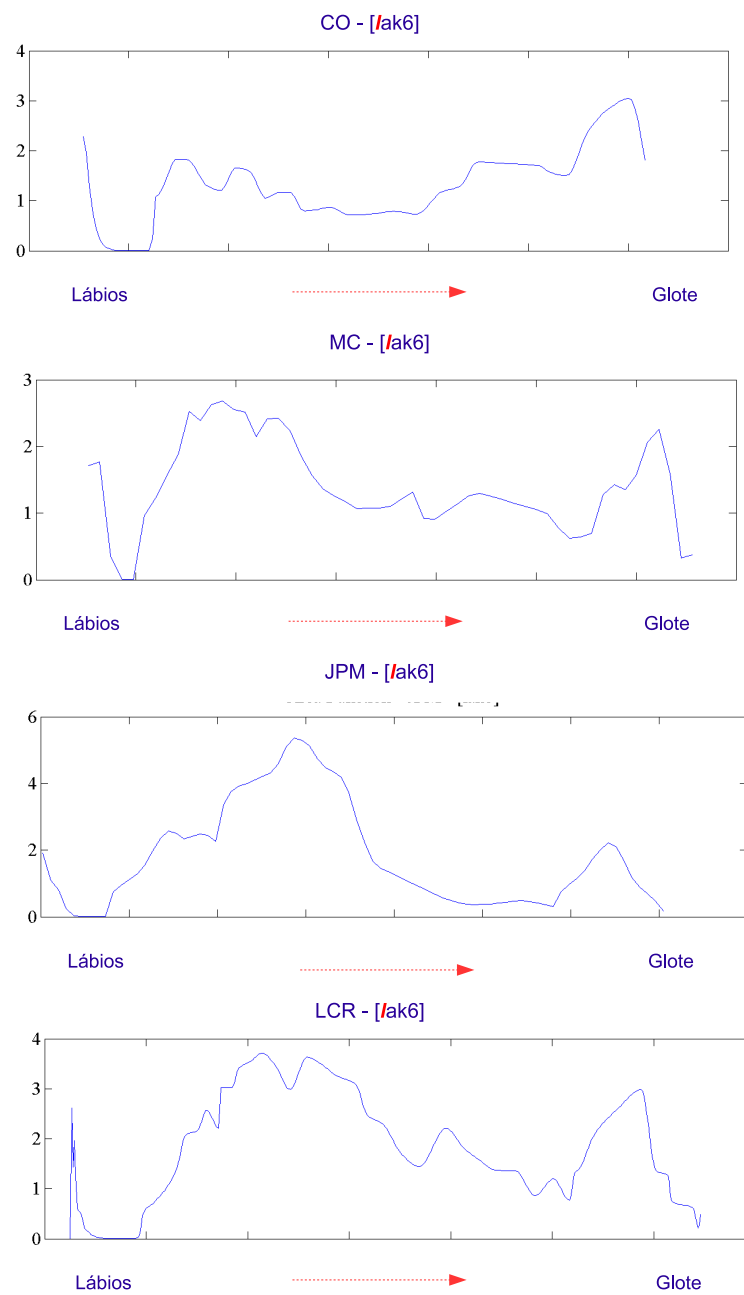


Figura 5.32: Exemplos de funções de área referentes ao /l/ para quatro falantes (CO, MC, JPM e LCR) na produção de [lak6].

visão global do padrão de variação de área para todos os falantes efetuou-se uma normalização do trato vocal (relativa ao comprimento e área) para cada um dos falantes e considerou-se depois a área normalizada média. Nesta situação, o valor absoluto das áreas não é considerado (a área passa a ser um valor relativo com intervalo de variação entre [0-1]).

A figura 5.33 apresenta-nos a função de área normalizada média (linha vermelha) para os seis falantes considerados para a produção de [lak6]. Os *boxplots* permitem ter uma ideia da variabilidade entre os falantes. Note-se que na zona do contacto alveolar e nas zonas uvular ou faríngea a variabilidade é menor.

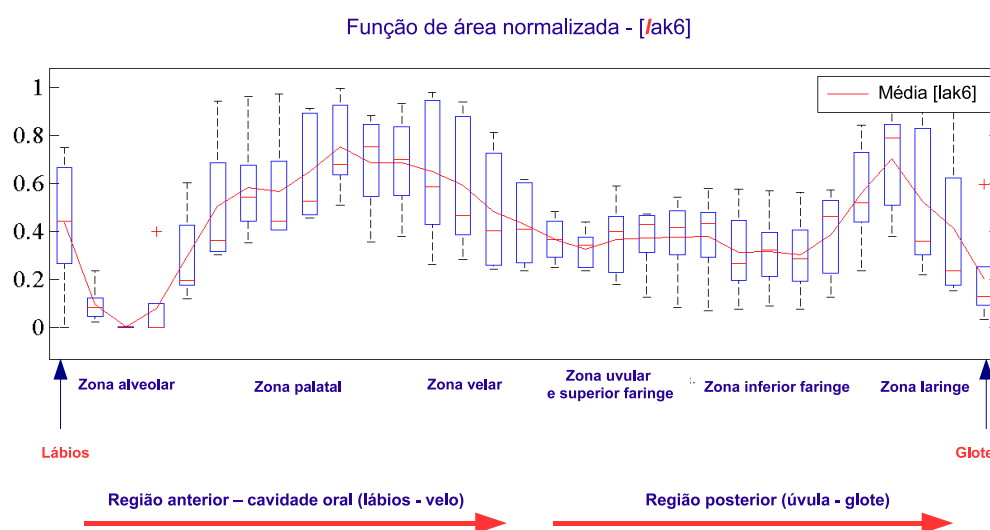


Figura 5.33: Função de área normalizada, para seis falantes, referente ao /l/ na produção de [lak6].

A figura 5.34, apresenta informação semelhante, para os seis falantes, para o /l/ produzido noutros contextos e posições silábicas. Em qualquer das imagens, pode-se constatar que as áreas mais elevadas se observam a nível da cavidade oral com valores máximos na zona média do palato. A partir daí as áreas diminuem em direção à zona do véu palatino e atingem os valores mais baixos na zona uvular e ao longo da faringe. Esta diminuição das áreas está associada ou a uma elevação do dorso da língua em direção à zona velar/uvular ou a um recuo em direção à faringe. Não há, como na zona alveolar, o estabelecimento de contacto, mas existe claramente uma segunda zona, ao longo do trato, em que se forma uma constrição.

5.5.1.4 Passagens laterais

O escape de ar lateralmente, através dos lados da língua, é uma das características dos sons laterais. A figura 5.35 apresenta algumas imagens, no plano coronal, em diferentes pontos do trato vocal, referentes ao falante JPM na produção de [lak6], com o intuito de ilustrar as descrições que efetuaremos relativas ao fluxo de ar lateral, às passagens ou canais laterais e cavidade supra-lingual. Para além disso, estão também assinaladas as zonas de contacto da língua na linha média

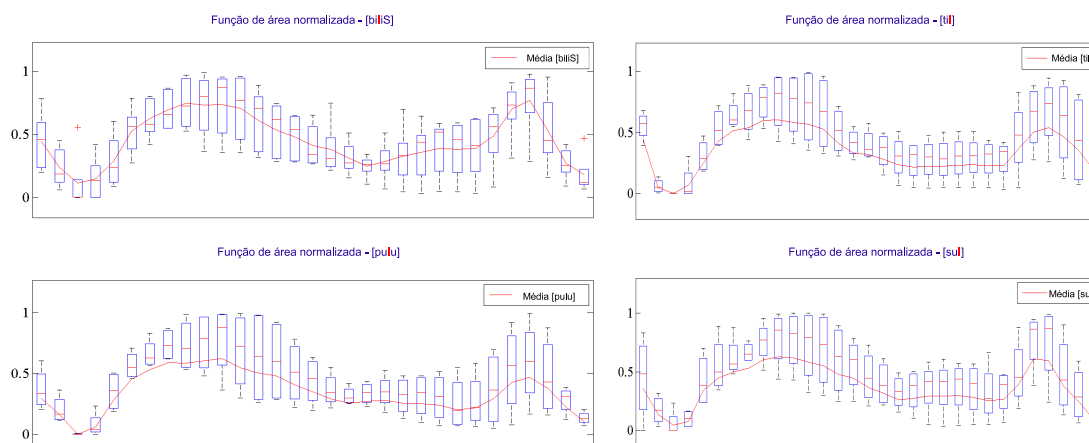


Figura 5.34: Funções de área normalizadas (para seis falantes) referentes ao /l/ em vários contextos e posições.

(zona dos alvéolos) e contactos laterais com o palato duro¹⁵.

Foi nosso objetivo obter dados relativos ao comprimento e área das passagens laterais. Convirá, neste momento, clarificar o que foi considerado como passagem lateral: apesar de o ar poder circular dos lados da língua, uma vez que ela se encontra comprimida lateralmente em direção à linha média, este fluxo de ar está conetado com o fluxo de ar central. Para as medidas efetuadas (comprimento e área), apenas foram tidas em consideração as áreas dos canais laterais quando estes se encontram separados do fluxo de ar central. Isso acontece, por exemplo, quando os lados da língua tocam lateralmente no palato duro, dando origem à formação de uma ou duas passagens do lado da língua e formação de uma terceira cavidade, separada, denominada cavidade supra-lingual. Como a língua não toca simetricamente no palato, existe alguma assimetria quer na extensão quer na área das passagens laterais. Noutros casos, só se estabelecem passagens laterais, quando a língua toca na zona alveolar, na linha média, e o ar se escapa lateralmente (i.e. dos lados do contacto alveolar).

As imagens apresentadas na figura 5.36 ilustram estas duas possibilidades. A figura mostra visualizações 3D dos cortes segmentados ao longo do trato vocal para dois falantes. Como se pode observar, a falante CO só apresenta passagens laterais na zona do contacto alveolar e não há formação de cavidade supra-lingual, enquanto que no falante LCR as passagens laterais se prolongam para trás do contacto dando origem à formação da cavidade supra-lingual.

A tabela 5.11 apresenta os valores relativos ao comprimento (ou extensão) das passagens laterais (esquerda e direita) e a área média observada, nas duas passagens laterais, para cada uma das produções do [l]. A informação é relativa a duas falantes femininas (CO e MC) e a dois falantes masculinos (JPM e LCR). As tabelas 5.12 e 5.13 apresentam, respetivamente, o valor médio do comprimento das passagens laterais por falante e o valor médio do comprimento de cada uma das

¹⁵Sobre estas imagens não foi efetuado qualquer tipo de processamento, trata-se apenas de visualização (*volume rendering*) e navegação através do volume adquirido.

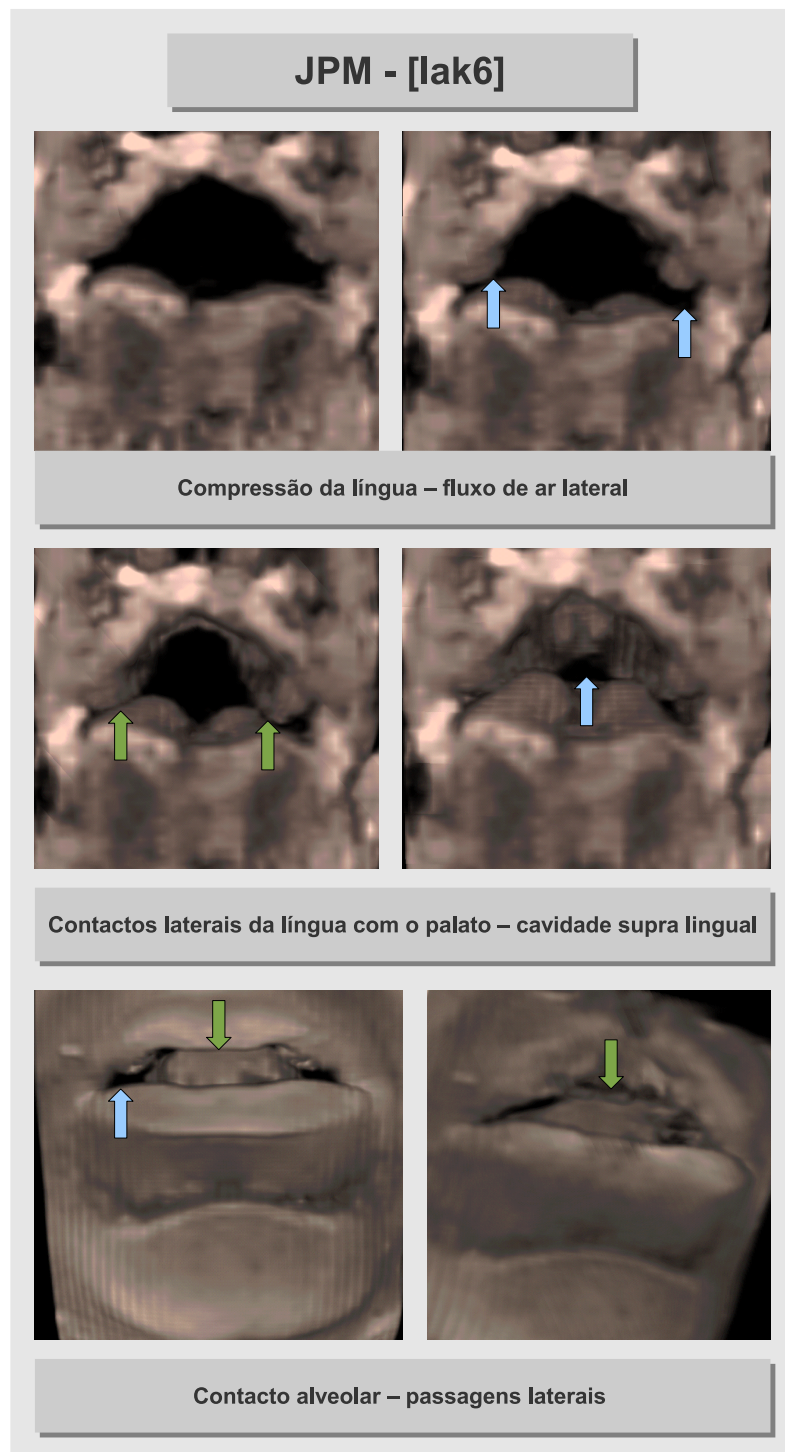


Figura 5.35: Imagens em diferentes zonas do trato vocal, do falante JPM, que mostram o fluxo de ar lateral, as passagens laterais, a cavidade supra-lingual e as zonas de contacto da língua com o palato lateralmente e na linha média.

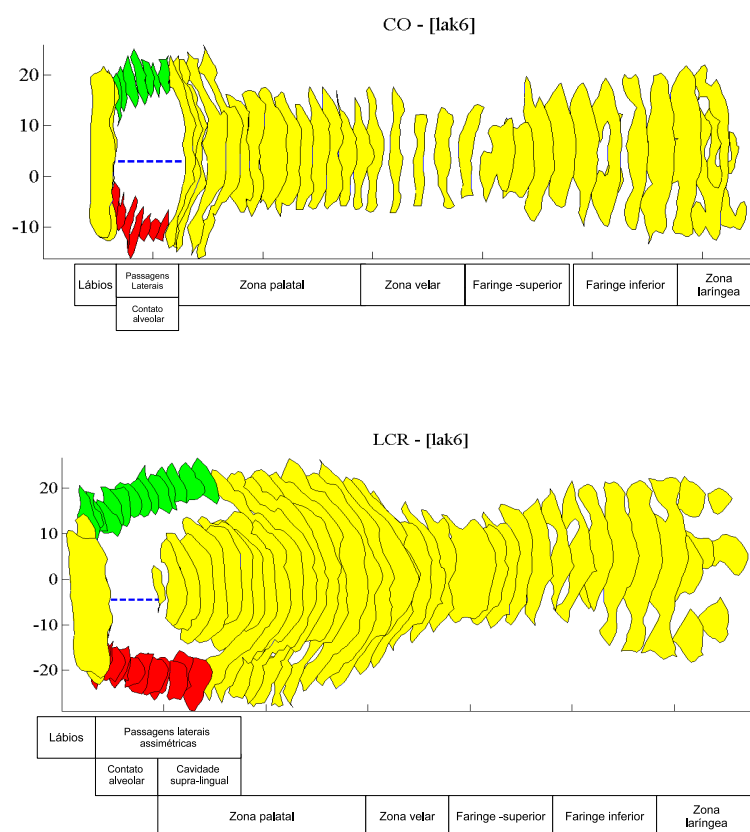


Figura 5.36: Exemplo de visualização 3D do trato vocal - série de cortes empilhados obtidos para o /l/ em posição inicial como em [lak6] para os falantes CO (em cima) e LCR (em baixo). É possível observar a zona de contacto alveolar, as passagens laterais e a cavidade supra-lingual separada (para o falante LCR). As restantes zonas do trato vocal estão identificadas.

passagens laterais em função da posição silábica (ataque, ataque intervocálico e coda).

Tabela 5.11: *Extensão das passagens laterais (mm) e área média das passagens laterais (cm²). Informação para duas falantes (CO e MC) e dois falantes (JPM e LCR).*

Falante	Palavra	PL esquerda (mm)	PL direita (mm)	Área média (cm ²)
CO	[lak6]	9.4	9.1	0.19
MC		2.7	4.8	0.25
JPM		27.8	32.3	0.21
LCR		19.1	23.0	0.36
CO	[sal]	4.8	4.2	0.19
MC		2.3	2.0	0.19
JPM		5.1	4.0	0.26
LCR		19.4	18.5	0.27
CO	[sal6]	4.2	4.8	0.22
MC		4.7	4.4	0.31
JPM		36.5	35.3	0.29
LCR		12.0	11.9	0.26
CO	[lit4u]	4.9	8.4	0.20
MC		5.7	4.8	0.21
JPM		44.1	43.8	0.29
LCR		17.5	20.9	0.24
CO	[til]	7.8	7.2	0.29
MC		3.3	5.7	0.19
JPM		20.9	23.7	0.19
LCR		18.9	16.7	0.19
CO	[biliS]	8.1	7.6	0.25
MC		2.4	2.3	0.26
JPM		18.4	9.7	0.20
LCR		15.4	18.4	0.21
CO	[lup6]	7.4	7.5	0.15
MC		5.4	4.5	0.22
JPM		23.3	28.3	0.16
LCR		10.3	9.3	0.22
CO	[sul]	6.8	6.3	0.27
MC		5.7	4.6	0.20
JPM		28.3	24.1	0.26
LCR		27.2	23.3	0.20
CO	[pulu]	3.1	2.5	0.28
MC		5.0	4.6	0.14
JPM		26.5	29.7	0.19
LCR		5.5	5.8	0.29

Considerando todas as produções efetuadas pelos falantes, pode observar-se na tabela 5.11 que a área média das passagens laterais variou entre 0.14 cm² (numa produção da falante MC) e 0.36 cm² para uma das produções do falante LCR. A área mínima observada foi de 0.05 cm² e a máxima de 0.56 cm² (dados não incluídos na tabela).

Tabela 5.12: *Extensão das passagens laterais em (mm), valor médio, máximo e mínimo, por falante.*

Género	Falante	PL (Média - mm)	Máximo (mm)	Mínimo (mm)	DP
F	CO (n = 18)	6.1	9.4	2.5	2.0
	MC	4.2	5.7	2.0	0.8
M	JPM	25.7	44.1	4.0	11.4
	LCR	16.3	27.2	5.5	6.0

A análise da tabela 5.12 permite constatar que a extensão das passagens laterais varia consideravelmente entre os falantes. A falante MC é a que apresenta passagens laterais mais curtas com um valor médio de 4.2 *mm* para todas as produções efetuadas. O falante que apresentou passagens laterais de maior dimensão foi o falante JPM (25.7 *mm*), sendo também neste falante que se registou uma maior diferença entre o valor mínimo e máximo (4.0–44.1 *mm*). As passagens laterais são, em média, mais compridas nos falantes masculinos (21.0 *mm*) do que nas falantes femininas (valor médio de 5.2 *mm*). Constatamos que apesar do comprimento das passagens ser menor, em média, nas falantes femininas, as áreas destas nem sempre o foram.

De acordo com a tabela 5.11, verifica-se que as passagens direita e esquerda são ligeiramente assimétricas, em comprimento, sendo que nas falantes femininas essa assimetria é menor (inferior ou próximo de 1 *mm* - com exceção de uma produção da falante CO) do que nos falantes masculinos.

Procurou-se ainda avaliar as diferenças nas passagens laterais em função da posição silábica. Assim, foram consideradas para a análise todas as produções de todos os falantes, para uma mesma posição silábica (cf. tabela 5.13). Verificamos que as passagens laterais são ligeiramente mais compridas (cerca de 4 *mm*) em posição de ataque do que nas posições de ataque intervocalico e de coda. Nestas duas últimas posições, os valores são similares. Todavia, analisando as produções de cada falante, individualmente, nem sempre se verificou este padrão (cf. tabela 5.11).

Tabela 5.13: *Extensão das passagens laterais (mm), valor médio para as passagens esquerda e direita, em função da posição silábica.*

Posição	PL (esquerda)	PL(direita)
Ataque (n=12)	16.4	16.2
At.inter	11.1	11.4
Coda	12.5	11.7

A tabela 5.14 apresenta os valores médios, por género, para cada uma das posições silábicas. A diferença entre a posição de ataque e as restantes posições é muito mais marcada para os falantes masculinos (cerca de 8 *mm*) do que para os femininos (cerca de 1 *mm*).

Tabela 5.14: *Extensão das passagens laterais em (mm), valor médio para as passagens esquerda e direita, em função da posição silábica, por género.*

Género	Posição	PL (esquerda)	PL (direita)
F	Ataque	5.6	6.2
	At.inter	4.6	4.4
	Coda	5.1	5.1
M	Ataque	27.2	26.3
	At. inter	19.1	18.5
	Coda	20.0	18.4

A tabela 5.15 apresenta os resultados da dimensão das passagens laterais em função do contexto vocálico. Consideraram-se as produções em posição de ataque. Verifica-se que, para qualquer dos géneros, as passagens laterais são um pouco maiores quando o [l] foi produzido no contexto da vogal [i]. No entanto, as diferenças, em função da vogal, não são dignas de nota.

Finalmente, as imagens da figura 5.37 mostram visualizações 3D do trato vocal com as passagens

laterais destacadas (vermelho e verde), para dois falantes (um masculino e outro feminino), para cada uma das três vogais, com o /l/ em posição inicial e final de sílaba.

Tabela 5.15: *Extensão das passagens laterais em (mm), valor médio para as passagens esquerda e direita, por gênero e vogal, para a posição de ataque.*

Gênero	Vogal	PL (esquerda)	PL (direita)
F	[a]	4.7	4.9
	[i]	5.4	6.0
	[u]	5.2	4.7
M	[a]	20.0	20.8
	[i]	22.6	22.2
	[u]	20.2	20.1

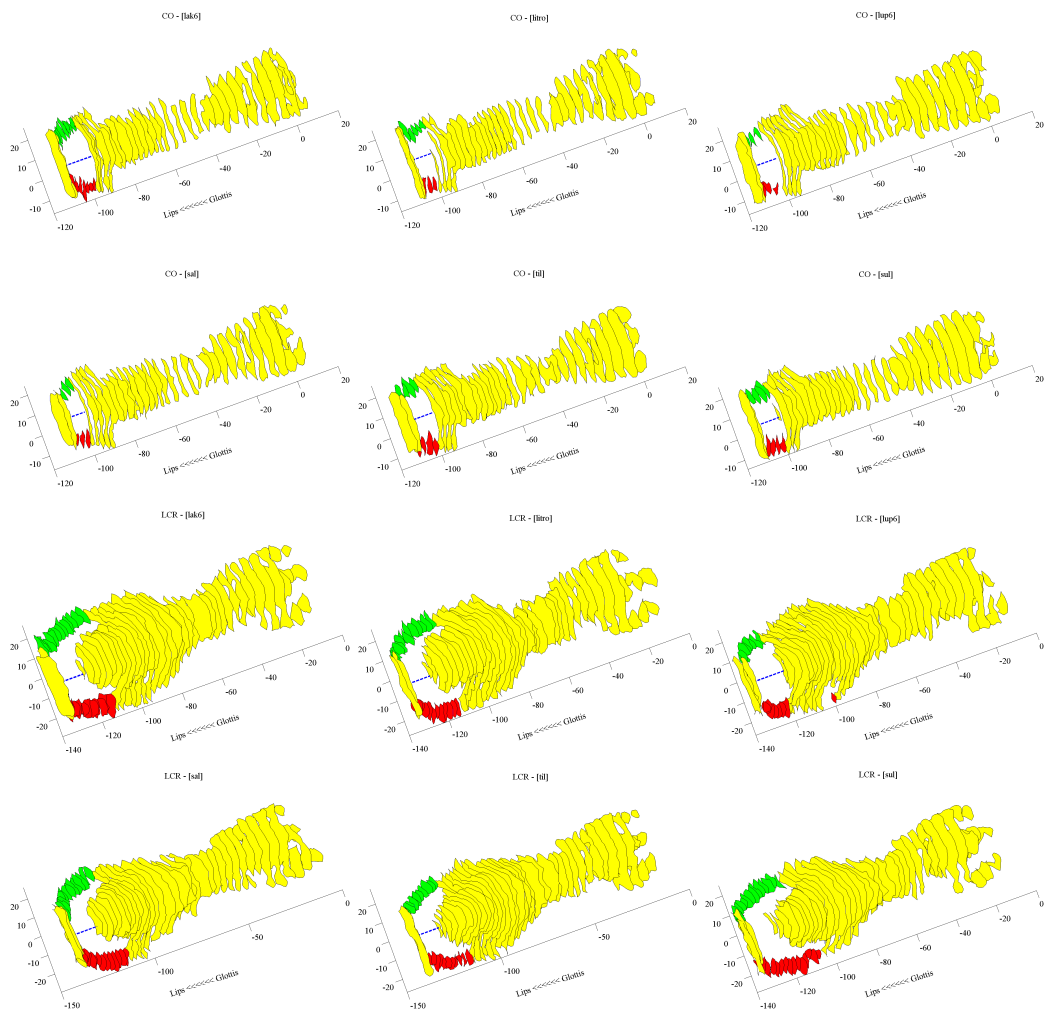


Figura 5.37: Visualizações 3D do trato vocal, ilustrando as passagens laterais, para dois falantes (CO e LCR), para o /l/ produzido em posição inicial e final, no contexto das vogais [a], [i] e [u]

5.5.1.5 Efeito da posição silábica (2D e 3D)

Após uma caracterização genérica das características, 2D e 3D, da língua e do trato vocal procuramos de seguida, tendo em vista os objetivos propostos, estabelecer a comparação das produções do /l/ nas três posições silábicas consideradas. Esta informação permitir-nos-á verificar se existem ou não diferenças na configuração articulatória do /l/ em função da posição silábica. Para estabelecer esta comparação, recorreremos à sobreposição dos contornos 2D que nos permitem obter informação qualitativa, aos valores dos IP (obtidos por comparação entre contornos) que possibilitam a quantificação das diferenças e, finalmente, às funções de área do trato vocal (informação 3D).

Assim, na figura 5.38, apresenta-se a sobreposição dos contornos do /l/, para seis falantes, em posição de ataque, ataque intervocálico e em coda, relativos às suas produções no contexto da vogal [a]. A figura 5.39 mostra, para a falante CO, os contornos sobrepostos para as três posições silábicas nos contextos vocálicos considerados [a,i,u]. A tabela 5.16 apresenta os IP's calculados para todos os pares de contornos, considerando as diferentes posições silábicas e vogais, para todos os falantes.

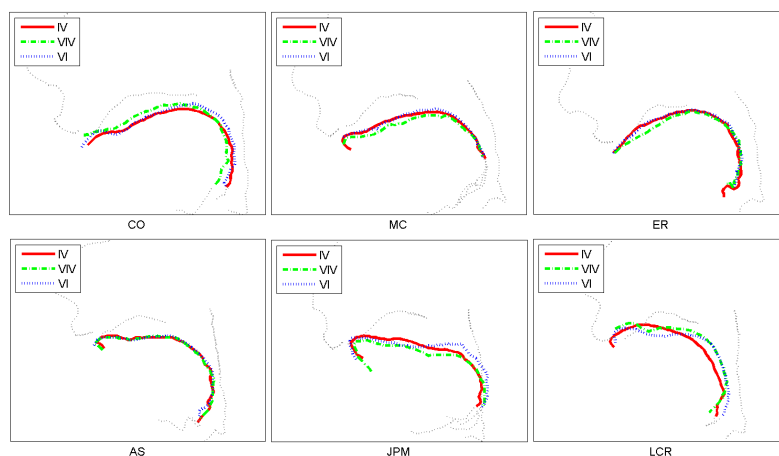


Figura 5.38: Sobreposição dos contornos 2D para seis falantes para o /l/, em posição de ataque (LV), ataque intervocálico (VLV) e coda (VL), no contexto da vogal [a]. Falantes femininas CO, MC, ER (em cima) e, em baixo, falantes masculinos (AS, JPM, LCR).

A análise através da sobreposição de contornos (cf. figura 5.38) permite-nos verificar que, no plano médio, a nível da ponta da língua (zona do contacto alveolar), não existem diferenças óbvias em função da posição silábica para qualquer um dos falantes. Para as falantes femininas (em cima), a zona posterior do dorso não apresenta variações, observando-se algumas diferenças na altura da parte média do dorso da língua, posterior ao contacto alveolar. O falante masculino (AS) não apresenta diferenças visíveis nos contornos para as três produções do /l/, sendo os falantes JPM e LCR os que apresentam maior variabilidade.

Os valores dos IP's obtidos para as diferentes posições consideradas corroboram estas observações. Note-se, por exemplo os valores mais elevados para o falante AS e os valores mais baixos (diferenças maiores) para os falantes JPM e LCR. Os valores dos IP's para as falantes

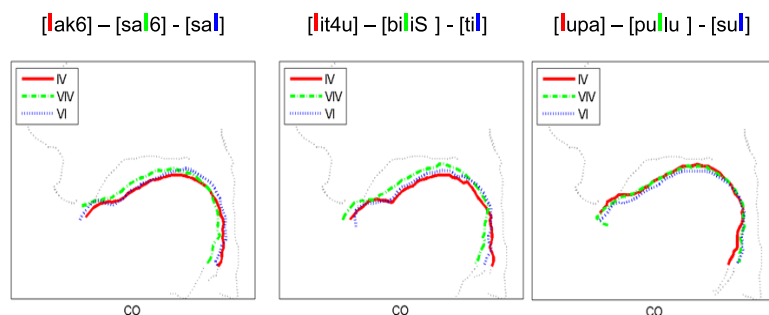


Figura 5.39: Efeito da posição silábica - Sobreposição dos contornos 2D para uma falante (CO) considerando produções do /l/ em posição de ataque (LV), ataque intervocálico (VLV) e coda (VL), no contexto das vogais [a] (esquerda), [i] (centro), [u] (direita).

femininas são também, na generalidade, elevados. Refira-se que a comparação através dos IP's reflete a diferença global entre dois contornos, não permitindo aferir diferenças em partes específicas da língua.

Tabela 5.16: Valores dos IP calculados para todos os pares de contornos, em função da posição silábica. *Pi-Pm* representa a comparação de contornos em posição de ataque (*Pi*) e ataque intervocálico (*Pm*), *Pi-Pf* representa a comparação de contornos entre a posição de ataque e de coda (*Pf*) e *Pm-Pf* a comparação dos contornos em posição de ataque intervocálico e coda.

Género	Falante	[a]			[i]			[u]		
		<i>Pi-Pm</i>	<i>Pi-Pf</i>	<i>Pm-Pf</i>	<i>Pi-Pm</i>	<i>Pi-Pf</i>	<i>Pm-Pf</i>	<i>Pi-Pm</i>	<i>Pi-Pf</i>	<i>Pm-Pf</i>
F	CO	0.621	0.776	0.631	0.504	0.791	0.624	0.850	0.706	0.808
	MC	0.710	0.895	0.633	0.399	0.735	0.538	0.488	0.624	0.698
	ER	0.765	0.889	0.774	0.754	0.761	0.728	0.405	0.564	0.722
M	AS	0.889	0.869	0.868	0.768	0.766	0.893	0.596	0.687	0.824
	JH	0.736	0.482	0.708	0.808	0.802	0.757	0.747	0.753	0.898
	JPM	0.539	0.538	0.522	0.397	0.539	0.469	0.514	0.482	0.757
	LCR	0.481	0.446	0.610	0.550	0.665	0.684	0.547	0.603	0.685
	Média	0.677	0.699	0.678	0.597	0.722	0.670	0.592	0.631	0.770

Para facilitar a análise dos resultados, elaboramos a partir dos dados da tabela, alguns gráficos de barras, em que consideramos de forma separada, o efeito da posição silábica para todos os falantes e a comparação dos contornos em posição inicial e final em função da vogal (cf. figura 5.40). Apresenta-se ainda o resultado da comparação de todos os contornos nas várias posições, por falante e por género (cf. figura 5.41).

A análise do gráfico da esquerda da figura 5.41 permite verificar que são os falantes JPM, LCR e ER que no conjunto de todas as produções apresentam maiores diferenças entre os pares de contornos (i.e. maior variabilidade). Os falantes MC, CO e AS apresentam menor variabilidade (com IP médio acima de 0.760). Efetuando uma separação, por género, para todas as produções (cf. gráfico da direita da figura 5.41) verifica-se que no conjunto há menos variabilidade (0.727) nas falantes do género feminino do que nos masculinos (0.694).

Interessa agora observar o que se passa quando passamos da informação 2D para 3D. As

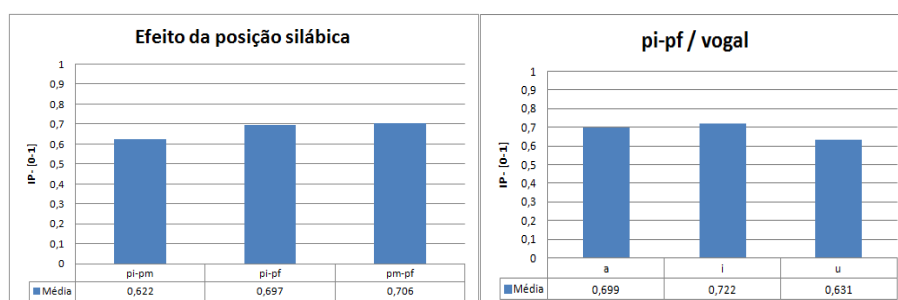


Figura 5.40: Média dos IP obtidos para todos os falantes, nas diferentes posições silábicas (esquerda). IP médios para o par ataque (pi) e coda (pf), em função da vogal em que foi produzido (direita).

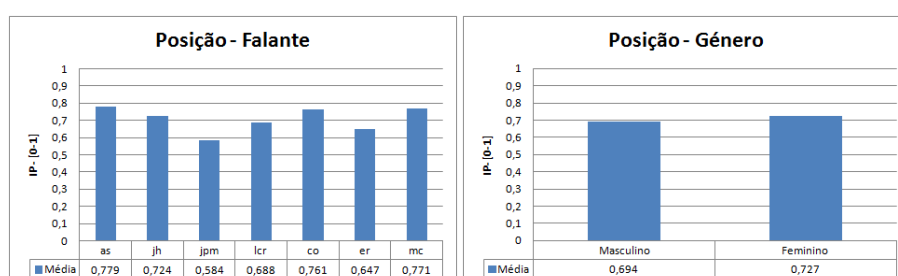


Figura 5.41: Média dos IP por falante (à esquerda) e por género (à direita), considerando todas as posições silábicas.

imagens que se apresentam nas figuras seguintes (figuras 5.42 a 5.45) apresentam sobreposição das funções de área do trato vocal para as diferentes posições silábicas. Para simplificar apresentam-se, apenas, comparações das funções de área obtidas para a posição de ataque e de coda, para quatro falantes (dois femininos e dois masculinos).

Relativamente à falante CO (ver figura 5.42), é evidente que não existem, nas produções em ataque e em coda, variações dignas de registo nas funções de área do /l/, independentemente do contexto vocálico em que foi produzido. No caso do [u], note-se que esta informação está perfeitamente de acordo com o observado em 2D. No caso da vogal [i], as diferenças observadas nos contornos 2D esbatem-se completamente em 3D.

A figura 5.43, apresenta informação para quatro falantes, considerando as funções de área em posição de ataque e de coda no contexto da vogal [a]. As figuras 5.44 e 5.45 apresentam a mesma informação para o /l/ produzido no contexto das vogais [i] e [u].

A análise das figuras em que se consideram as produções dos quatro falantes permite verificar que as falantes femininas apresentam menos diferenças nas produções (ataque/coda) do que os falantes masculinos. Note-se no entanto que, no caso dos falantes masculinos, na zona da faringe as áreas são baixas para as duas produções. Nas falantes femininas, as áreas são menores na zona correspondente à área do véu palatino e da úvula. Para todos, há contacto na zona alveolar com pequenas diferenças na extensão do mesmo.

Uma vez que é complexo estabelecer comparações entre todos os falantes, em virtude as já referidas diferenças de comprimento do trato e respetivas áreas, recorreremos, de novo, às funções

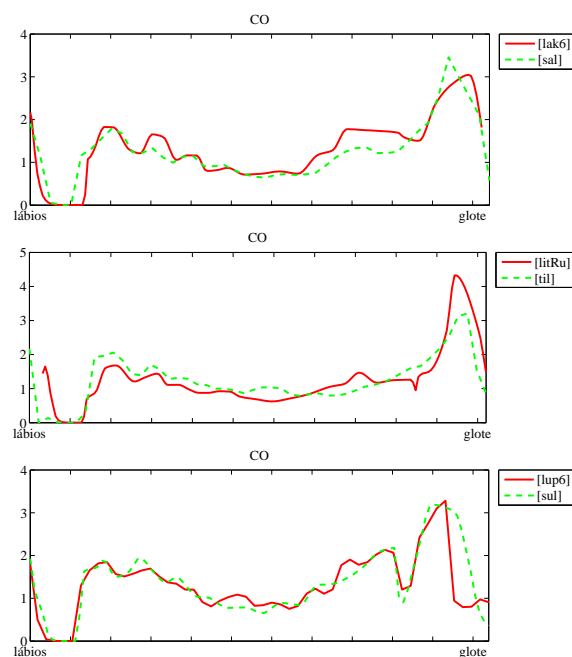


Figura 5.42: Funções de área do trato vocal para o /l/, em posição de ataque ([lak6]) e de coda ([sal]), para cada vogal [a, i, u] referentes à falante CO.

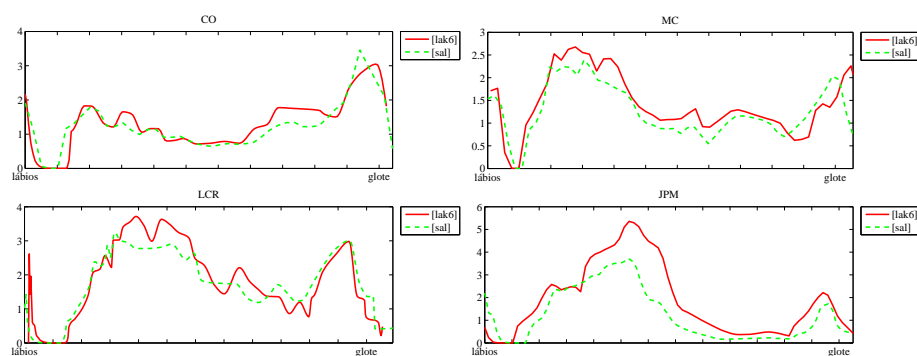


Figura 5.43: Funções de área do trato vocal para o /l/, em posição de ataque (vermelho) e de coda (verde), no contexto da vogal [a] referentes aos falantes CO, MC, LCR e JPM.

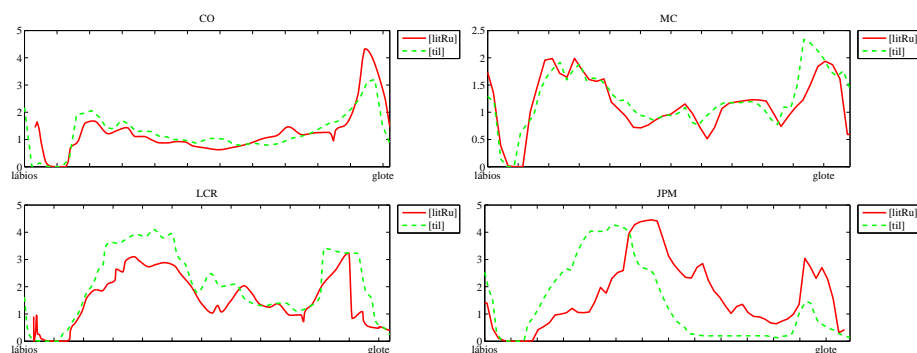


Figura 5.44: Funções de área do trato vocal para o /l/, em posição de ataque (vermelho) e de coda (verde), no contexto da vogal [i] referentes aos falantes CO, MC, LCR e JPM.

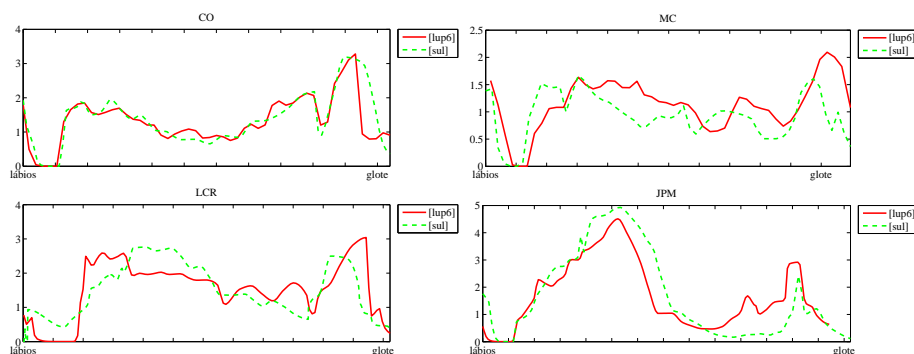


Figura 5.45: Funções de área do trato vocal para o /l/, em posição de ataque (vermelho) e de coda (verde), no contexto da vogal [u] referentes aos falantes CO, MC, LCR e JPM.

de área normalizadas para proporcionar uma visão global do comportamento das áreas ao longo do trato para todos os falantes. A figura 5.46 apresenta a área normalizada média, na produção de [lak6] (em cima) e [sal] (em baixo). Como se pode observar, na zona da laringe as áreas são relativamente elevadas e decrescem na zona da faringe e do véu palatino; aumentam depois até meio da zona palatal e começam a decrescer para a zona do contacto alveolar. Note-se que onde existe menor variabilidade é, precisamente, na zona alveolar, na zona do véu palatino e/ou superior da faringe.

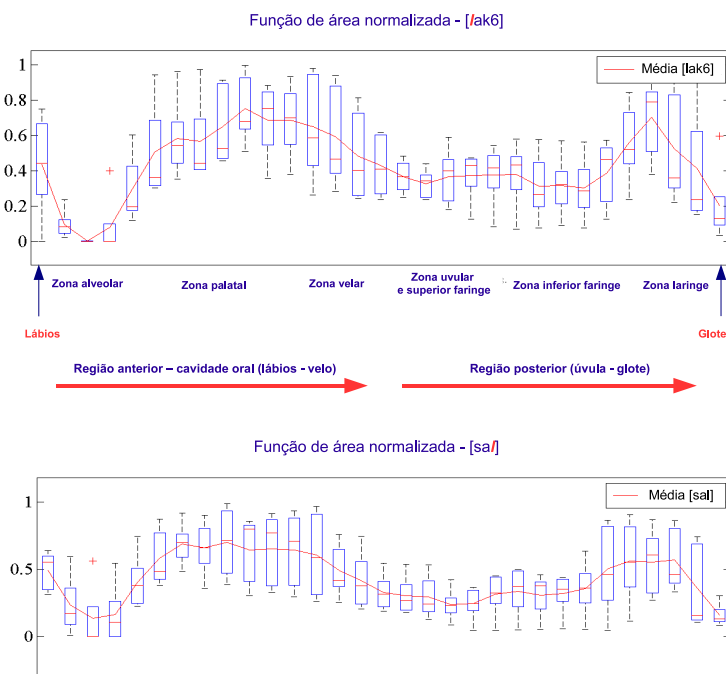


Figura 5.46: Exemplo de função de área normalizada, para seis falantes, na produção de [lak6], em cima e [sal] em baixo. A curva representa o valor médio da área. As diferentes zonas do trato são identificadas na imagem referente a [lak6].

A figura 5.47 apresenta informação semelhante considerando agora o /l/ produzido em posição de ataque e coda no contexto da vogal [u]. O padrão de variação das áreas é semelhante ao observado para o /l/ no contexto da vogal [a].

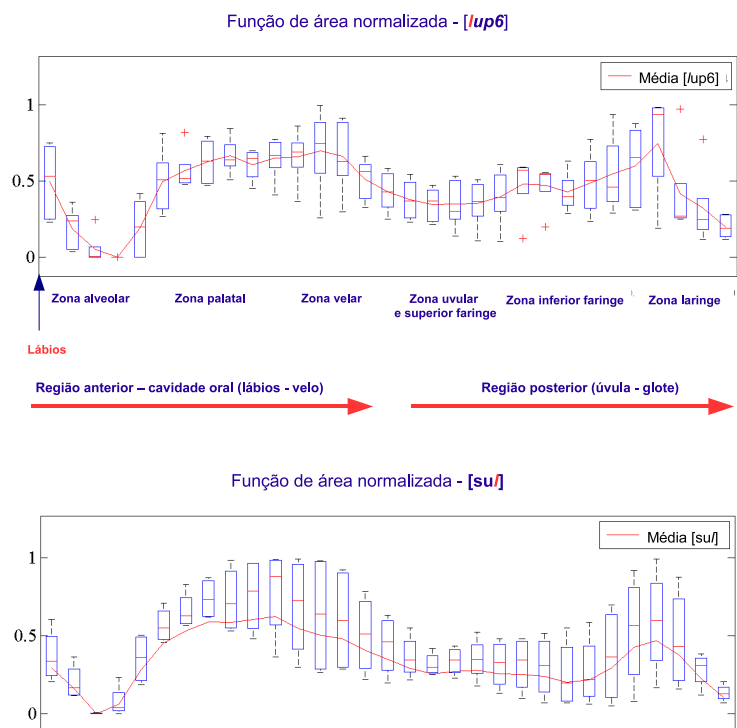


Figura 5.47: Exemplo de função de área normalizada, para seis falantes, na produção de [lup6], em cima e [sul] em baixo. A curva representa o valor médio da área. As diferentes zonas do trato são identificadas na imagem referente a [lup6].

No global, apesar da variabilidade e diferenças que possam existir entre os falantes, verifica-se que o padrão de variação das áreas entre as produções do /l/ efetuadas em ataque e coda é semelhante. Em alguns falantes há um maior recuo da língua em direção à faringe com áreas mais baixas nesta zona (geralmente, um pouco maiores em ataque do que em coda). Noutros, ao invés de um recuo da língua para a faringe, existe uma elevação do dorso da língua em direção ao véu palatino que se verifica quer nas produções em ataque quer em coda.

Para encerrar esta secção, apresentamos apenas um exemplo (veja-se figura 5.48) de comparação das produções do /l/, em posição de ataque e coda e que ilustra a existência de diferenças entre os falantes. Os falantes representados foram os que apresentaram, globalmente, maior (JPM) e menor variabilidade (CO) nas diferentes produções. A figura mostra, para além das áreas do trato vocal, a sobreposição de malhas poligonais da língua. A sobreposição das malhas foi efetuada com recurso ao PolyMeco e pode ser um método alternativo (tendo por base a forma da língua) para comparar diferentes produções dos falantes. De acordo com a escala, as zonas a vermelho correspondem às maiores diferenças entre as malhas e a azul as menores diferenças. A informação relativa à língua (sobreposição das malhas) e do trato vocal (sobreposição

das funções de área) está em perfeita consonância. No caso do falante JPM a diferença nas áreas, entre a posição de ataque e coda, verifica-se na zona velar e superior da faringe o que corresponde precisamente às zonas da língua onde as diferenças são também maiores. Na falante CO a informação obtida através das malhas da língua indica que não existem diferenças relevantes e o mesmo se pode constatar a partir da sobreposição das funções de área.

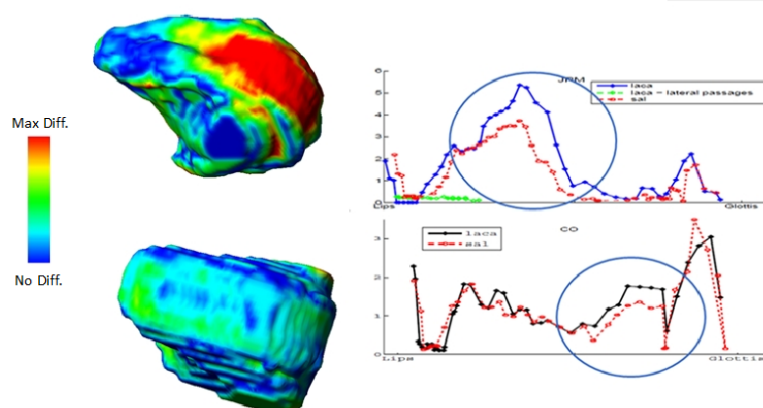


Figura 5.48: Variabilidade entre falantes, na produção do /l/ em ataque e coda, observada através da sobreposição das malhas 3D da língua (à esquerda) e respetivas funções de área do trato vocal (à direita), para os falantes JPM (em cima) e CO (em baixo).

5.5.1.6 Efeito do contexto vocálico

Foi também considerada a influência do contexto vocálico na produção do /l/. A figura 5.49 apresenta a sobreposição dos contornos 2D da língua, extraídos das imagens no plano médio sagital, nas produções referentes ao /l/, em posição de ataque, no contexto das vogais [a, i, u].

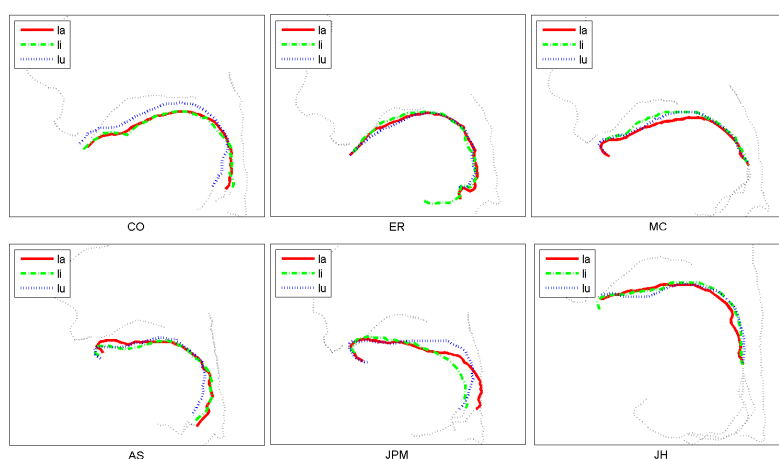


Figura 5.49: Sobreposição dos contornos 2D da língua, extraídos das imagens no plano médio sagital, nas produções referentes ao /l/, em posição de ataque, no contexto das vogais [a, i, u].

De forma similar ao efetuado na secção anterior, foram agora calculados os IP médios para os pares de contornos [a-i], [i-u] e [a-u] nas várias posições silábicas. Os resultados apresentam-se

na tabela 5.17.

Tabela 5.17: Valores dos IP calculados para os pares de contornos [a-i], [a-u] e [i-u], para cada uma das posições silábicas.

Género	Falante	<i>pi</i>				<i>pm</i>				<i>pf</i>			
		<i>a-i</i>	<i>a-u</i>	<i>i-u</i>	<i>Média</i>	<i>a-i</i>	<i>a-u</i>	<i>i-u</i>	<i>Média</i>	<i>a-i</i>	<i>a-u</i>	<i>i-u</i>	<i>Média</i>
Fem.	CO	0.877	0.480	0.447	0.602	0.889	0.796	0.864	0.849	0.855	0.792	0.855	0.834
	MC	0.671	0.777	0.867	0.771	0.874	0.725	0.683	0.761	0.866	0.770	0.707	0.781
	ER	0.837	0.871	0.787	0.832	0.742	0.535	0.398	0.558	0.714	0.577	0.386	0.559
Mas.	AS	0.730	0.552	0.635	0.639	0.810	0.818	0.769	0.799	0.899	0.872	0.932	0.901
	JH	0.818	0.723	0.848	0.797	0.758	0.790	0.664	0.737	0.768	0.554	0.598	0.640
	JPM	0.570	0.529	0.493	0.531	0.795	0.620	0.679	0.698	0.498	0.588	0.489	0.525
	LCR	0.673	0.565	0.630	0.623	0.644	0.720	0.812	0.725	0.772	0.726	0.652	0.717
	<i>Média</i>	0.740	0.642	0.672	0.684	0.787	0.715	0.696	0.733	0.767	0.697	0.660	0.708

Considerando agora a informação conjunta obtida através da sobreposição dos contornos (posição inicial) e os valores dos IP para a mesma posição (coluna da esquerda da tabela 5.17) verificamos que são os falantes MC, JH e ER que apresentam menos variabilidade (IP entre 0.771 a 0.832) nas produções efetuadas no contexto das três vogais. O falante JPM é de todos os falantes o que apresenta mais variabilidade (0.531). A tabela permite verificar que a média dos IP para todos os pares de contornos é mais elevada (menos influência da vogal) em posição intervocálica (0.733) do que nas outras duas posições. A maior variabilidade é observada em posição inicial com um IP médio mais baixo (0.684). No entanto, a diferença entre os vários IP's não será muito relevante. Poder-se-á dizer que os efeitos da vogal na produção do /l/ serão equivalentes nas três posições silábicas. Para qualquer das posições consideradas verifica-se uma maior semelhança nos contornos do [a] com [i] do que entre [a-u] ou [i-u].

Para averiguar os efeitos da vogal em 3D, efetuou-se a sobreposição das funções de área para cada falante para as produções do /l/ no contexto das vogais [a, i, u]. A figura 5.50 apresenta uma exemplo para dois falantes, um feminino (CO) e outro masculino (LCR).

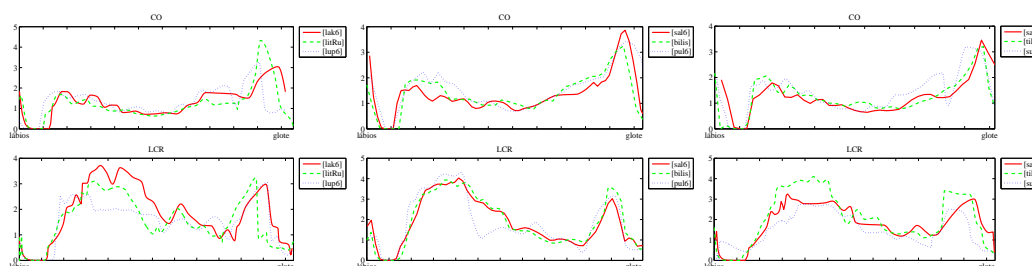


Figura 5.50: Efeito do contexto - Funções de área do trato vocal para o /l/, produzido no contexto das vogais [a] [i] e [u] nas três posições silábicas consideradas para dois falantes.

Neste caso, verifica-se que, em qualquer das posições silábicas, a falante CO apresenta menos variação nas curvas obtidas no contexto das três vogais. No falante LCR, há alguma variação nas áreas do trato vocal, com o contexto vocálico, mas essa variação é observada sobretudo a nível da parte média e posterior do dorso da língua. Note-se que, na zona do contacto alveolar, imediatamente atrás deste e na zona da faringe não há variações nas áreas para as diferentes vogais.

Para este falante, há menor efeito vocálico na produção da consoante em posição intervocálica e efeito semelhante em posição inicial e final.

As figuras 5.51 e 5.52 apresentam informação para quatro falantes, considerando agora apenas as funções de área do /l/ no contexto das três vogais em posição inicial (ataque) e em posição final (coda).

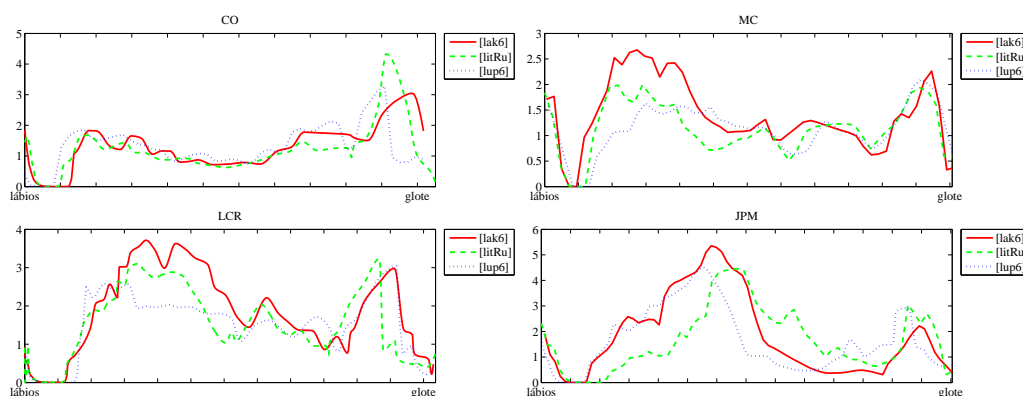


Figura 5.51: Efeito do contexto - Funções de área do trato vocal para o /l/, produzido no contexto das vogais [a] [i] e [u], em posição inicial para 4 falantes.

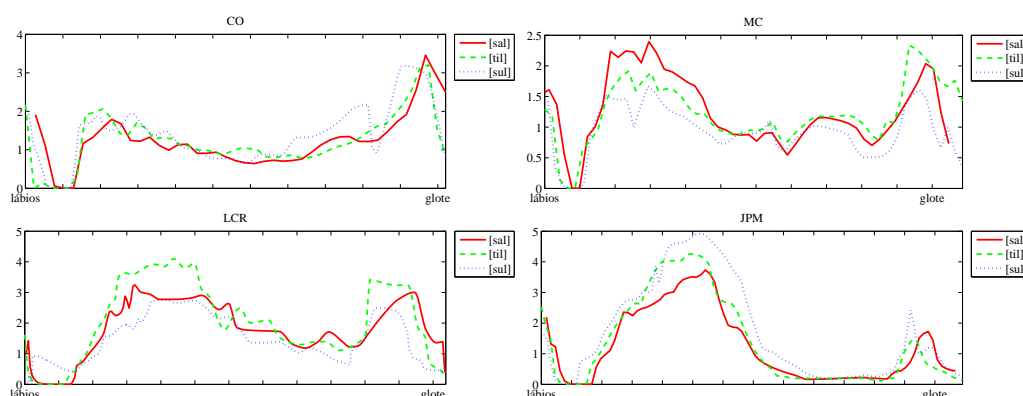


Figura 5.52: Efeito do contexto - Funções de área do trato vocal para o /l/, produzido no contexto das vogais [a] [i] e [u], em posição final para 4 falantes.

Para além das funções de área apresentamos também, a título de exemplo, como se pode obter informação relativa a efeitos coarticulatórios através da sobreposição das malhas superficiais da língua. A figura 5.53 apresenta a sobreposição e comparação de malhas 3D da língua, efetuada com recurso ao PolyMeco, para avaliar efeitos coarticulatórios na produção do /l/ em ataque (à esquerda) e em coda (à direita), para a falante CO. A informação está em linha com a informação 2D e com a fornecida através da sobreposição das funções de área, para esta falante. Considerando a escala colorida podemos constatar que não existem diferenças dignas de nota entre as malhas do /l/ produzido em posição inicial. Em posição final as diferenças são ainda mais subtis.

Uma análise global da informação permite-nos afirmar que há alguma influência da vogal na produção do /l/, particularmente, atrás do contacto alveolar, na zona média do dorso da língua.

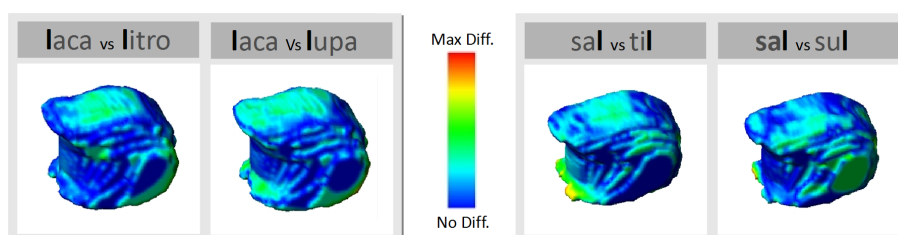


Figura 5.53: Efeito contextual em 3D - sobreposição e comparação de malhas 3D da língua, efetuada com recurso ao PolyMeco, para avaliar efeitos coarticulatórios na produção do /l/ em ataque (à esquerda) e em coda (à direita) para a falante CO. Imagem publicada em Martins *et al.* (2012a).

Estas diferenças são, no entanto, mais expressivas nos falantes masculinos. De todos os falantes o JPM é o que apresenta mais variabilidade. Os efeitos da vogal são equivalentes nas várias posições consideradas mas o /l/ produzido em posição inicial é ligeiramente mais afetado pela vogal que os produzidos em posição final e intervocálica.

5.5.2 A lateral palatal - /L/

Nas secções seguintes, apresentamos os resultados obtidos relativos ao /L/. Seguir-se-á uma estrutura semelhante à efetuada para o /l/, contemplando ponto e extensão da articulação, forma da língua, áreas ao longo do trato vocal, passagens laterais e efeitos coarticulatórios, contando com informação 2D e 3D. No final, reportam-se os resultados obtidos com uma primeira experiência de simulação efetuada para esta consoante.

5.5.2.1 Ponto e extensão do contacto linguo-palatal

A informação relativa ao ponto de articulação da consoante proveio essencialmente da análise da informação 2D, que se exemplifica através da figura 5.54. Na figura são apresentadas imagens obtidas no plano médio sagital, para o /L/, para seis falantes, no contexto da vogal [a]. Como se pode observar, todos os falantes representados na figura produzem a lateral com contacto da lâmina lingual e pré-dorso na região alveolar e pós-alveolar (ou pré-palatal). A falante ER (veja-se figura 5.55) é a única que articula a consoante com um maior envolvimento da região médio-dorsal da língua, sendo que o contacto contempla a zona pós-alveolar, estendendo-se à zona médio-palatal. Em alguns dos falantes, a articulação da palatal estende-se, até à zona dental (cf. imagens dos falantes LCR, JH e AS). A ponta da língua está baixa, para todos os falantes, quase tocando os incisivos inferiores.

Relativamente ao comprimento do contacto da língua com o palato na linha média (cf. tabela 5.18), verifica-se que este é, no geral, bastante maior do que o observado para o /l/. A extensão do contacto lingual, para todos os falantes e contextos, varia entre os 7.4 mm (falante JPM, contexto de [i]) e os 24.0 mm (falantes CO e ER, em contexto de [a]). Tendo em consideração as diferenças entre géneros no comprimento total do trato vocal (veja-se a secção 5.5.3), a extensão dos contactos é, proporcionalmente, maior nas mulheres do que nos homens. Considere-se, por

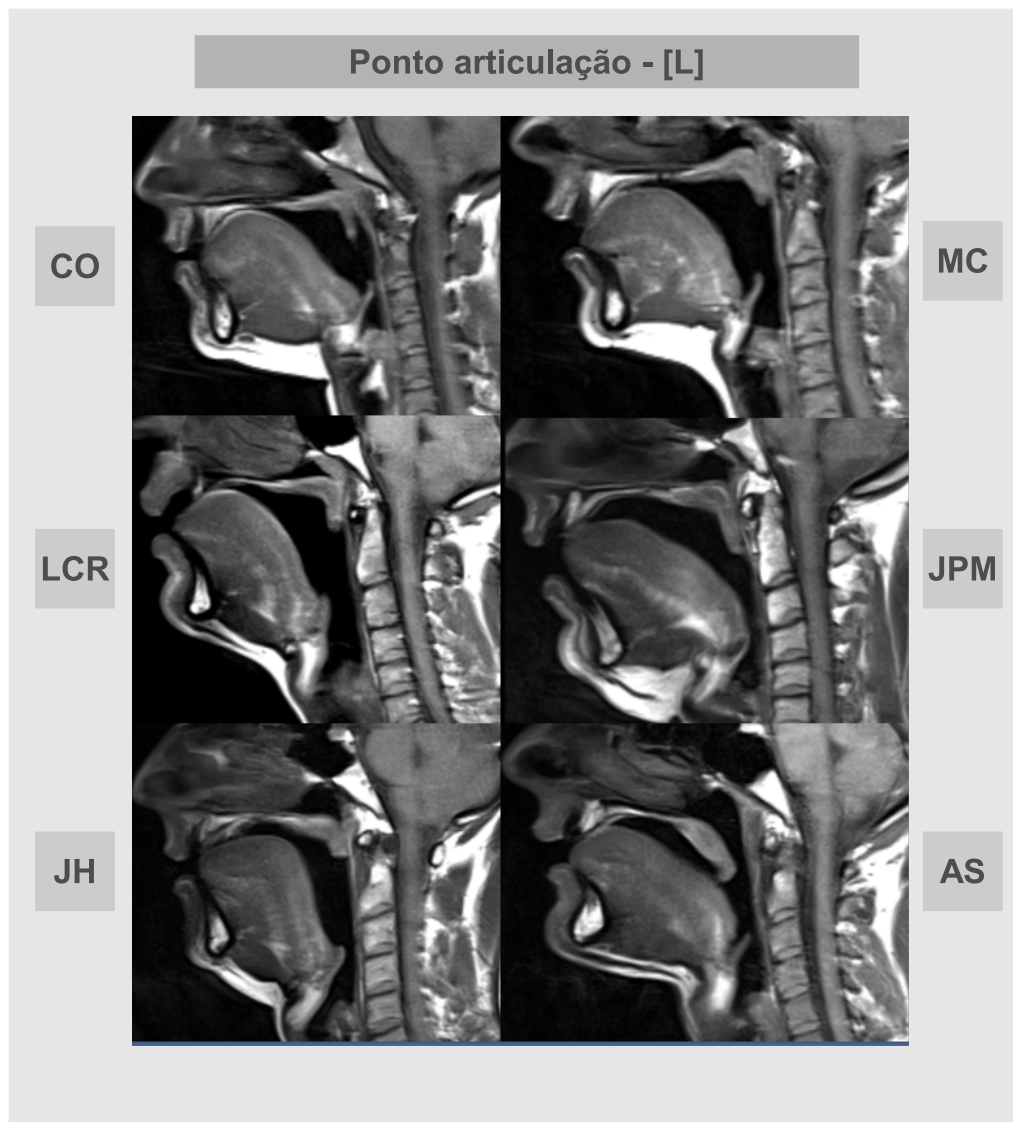


Figura 5.54: Ponto de articulação do [L] para seis falantes.



Figura 5.55: Ponto de articulação do [L] para a falante ER.

exemplo, o caso da falante CO que tem um trato vocal cerca de 3.6 *cm* mais curto que o falante JPM, mas a extensão do contacto na produção do /L/ é 3.0 *mm* maior, para o mesmo contexto.

Tabela 5.18: *Extensão dos contactos (em mm) na linha média referentes ao [L], para os 7 falantes.*

	CO (F)	MC (F)	ER (F)	JH (M)	LCR (M)	JPM (M)	AS (M)
[a]	24.1	10.3	24.0	9.2	15.3	21.1	15.9
[i]	10.3	11.1	21.0	10.7	19.4	7.4	20.0
[u]	21.2	14.4	-	8.1	10.1	18.2	14.9

5.5.2.2 Configuração da língua 2D e 3D

As imagens da figura 5.54 servem-nos também para verificar a configuração da língua em 2D. Contrariamente ao observado para o /l/, o dorso da língua apresenta-se bastante anteriorizado e numa posição elevada, particularmente na zona do pré-dorso. Nas imagens observa-se ainda que o /L/ foi produzido com o véu palatino numa posição elevada e fechado para todos os falantes, excetuando o falante AS, em que este se encontra aberto.

No que diz respeito à configuração da língua em 3D, tal como se pode observar nas malhas apresentadas na figura 5.56 (em baixo), verifica-se que esta partilha algumas características com a lateral alveolar. De facto, também para o /L/, a língua apresenta compressão lateral em direção à linha média e forma convexa do seu dorso. Esta compressão é, todavia, maior do que o observado para o /l/, estendendo-se à porção mais posterior do dorso, o que poderá justificar as áreas elevadas na zona velar observadas para o /L/ (ver secção seguinte). Note-se também a formação de *groove* na zona posterior da língua (a nível da raiz da língua). Contrariamente ao /l/, o ápice da língua está baixo e o dorso da língua está numa posição mais elevada do que para o /l/ como já havia sido observado nas imagens 2D.

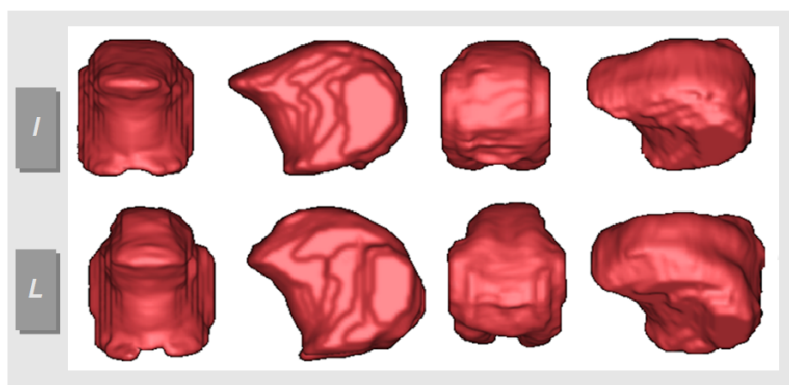


Figura 5.56: Visualizações 3D da língua para o /I/ (em cima) e para o /L/ (em baixo), relativas à falante CO, no contexto da vogal [a]. Da esquerda para a direita: vista anterior, lateral, posterior e superior.

5.5.2.3 Funções de área do trato vocal e passagens laterais

A análise das imagens 2D deixa já antever a dimensão da cavidade oral e da cavidade faríngea na produção do /L/. Os dados 3D, permitem, no entanto, a obtenção de medidas mais fidedignas. As figuras 5.57 e 5.58 apresentam as áreas segmentadas ao longo do trato vocal durante a produção do [L]. São visíveis as passagens laterais e a formação de uma cavidade supra-lingual separada.

A observação das imagens permite-nos ainda verificar que logo a partir da região velar a parte posterior e média do dorso da língua apresenta forma convexa e compressão lateral em direção à linha média, o que permite que o ar circule, lateralmente, dos dois lados da língua, contribuindo para as áreas relativamente elevadas que se observam nesta região. À medida que a parte anterior do dorso da língua se eleva, os bordos laterais da língua contactam com o palato lateralmente, permitindo que o ar passe por cima desta centralmente e se escape lateralmente, por entre os dentes. Formam-se, nesta zona, duas passagens laterais separadas e a cavidade supra-lingual, de forma semelhante ao que foi verificado, em alguns falantes, na produção da lateral alveolar. A cavidade supra-lingual estende-se até à zona do contacto que se estabelece na zona alveolar e pós-alveolar, na linha média. Na zona do contacto, cessa o fluxo de ar central sendo a passagem do ar efetuada, exclusivamente, pelos lados da língua. Na produção da palatal, houve sempre formação da cavidade supra-lingual e as passagens laterais estendem-se para lá da zona de contacto, sendo maiores de que as observadas na produção da lateral alveolar.

A figura 5.59 mostra as funções de área do trato vocal (azul tracejado) para duas falantes femininas (CO e MC) e dois falantes masculinos (JPM e LCR) relativos ao [L]. São também representadas as funções de área relativas às passagens laterais direita e esquerda (vermelho e verde), a cavidade supra lingual (preto) e a cavidade frontal anterior à constricção (azul). A tabela 5.19 inclui medidas do comprimento e as áreas médias das diferentes cavidades consideradas. Nas funções de área apresentadas na figura 5.59, é possível constatar a existência de uma diminuição progressiva das áreas ao longo da cavidade oral (direção glote-lábios) relacionada com a anteriorização e elevação da parte anterior do dorso da língua. Pode-se também observar que as passagens laterais se iniciam antes da zona de contacto, estendem-se lateralmente a esta e são ligeiramente

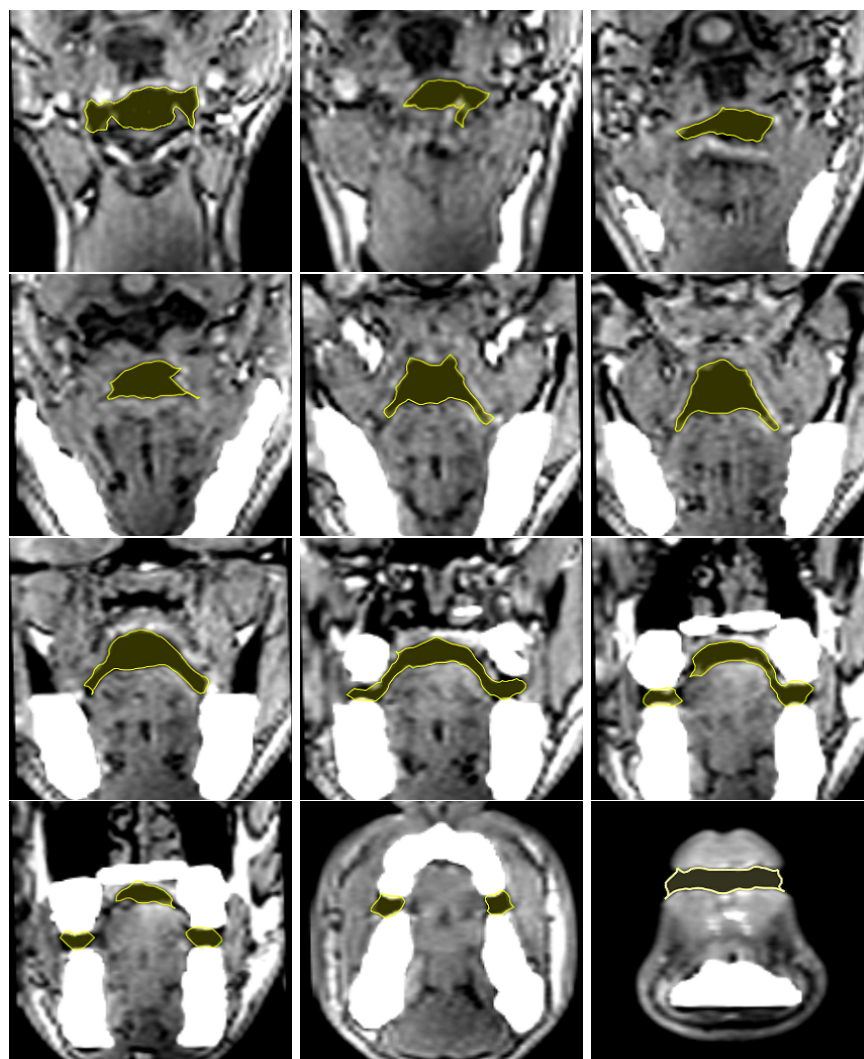


Figura 5.57: Áreas segmentadas ao longo do trato vocal para o falante LCR, na produção de [paL6].

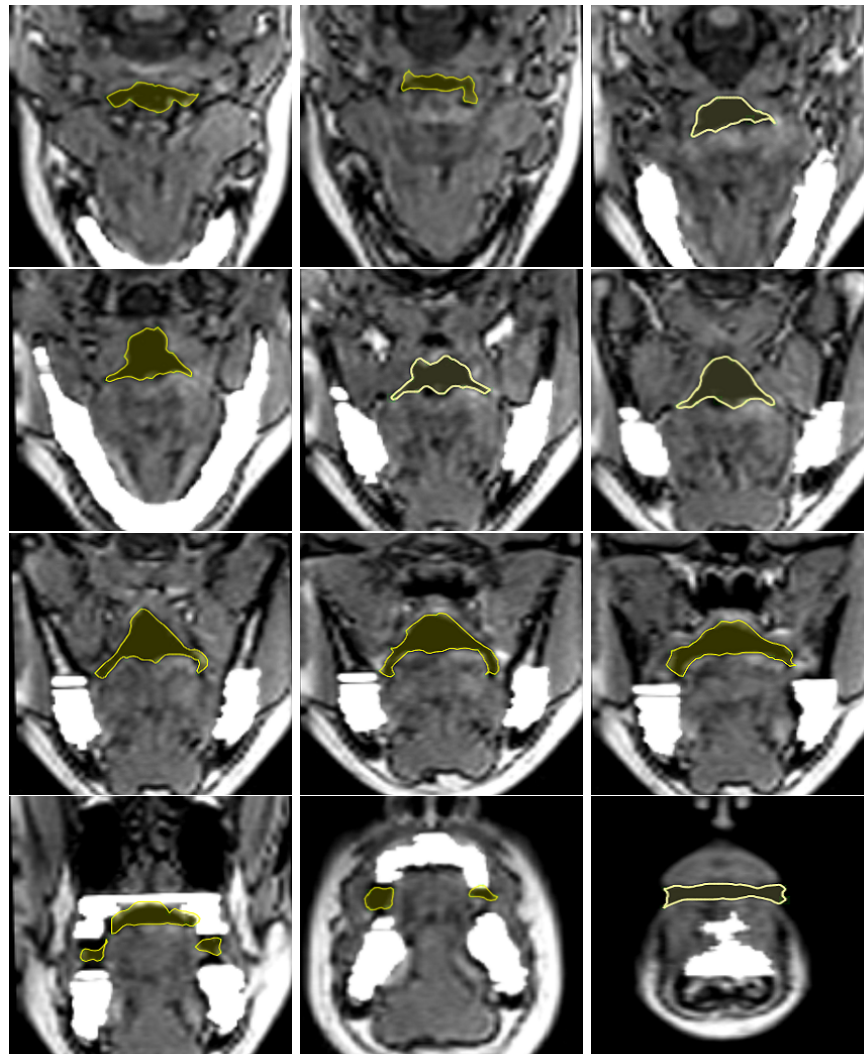


Figura 5.58: Áreas segmentadas ao longo do trato vocal para a falante MC, na produção de [paL6].

assimétricas em comprimento e em área (cf. tabela 5.19). Apenas no caso da falante CO, as áreas de cada uma das passagens laterais (direita e esquerda) são equivalentes. Nos falantes MC, JPM e LCR, a assimetria nas áreas é mais marcada. Esta assimetria nas áreas deve-se ao facto da língua se encostar mais a um lado do que ao outro dentro da cavidade oral. As áreas médias das passagens laterais foram, na generalidade, maiores do que as obtidas para a lateral alveolar (cf. tabela 5.11 para os mesmos falantes).

O comprimento da cavidade supra-lingual situou-se entre os 3.0–3.9 *cm*, com áreas médias que variaram entre 0.38–1.44 *cm*². Considerando as diferenças de comprimento do trato vocal entre géneros, a cavidade supra-lingual é proporcionalmente maior nas mulheres do que nos homens.

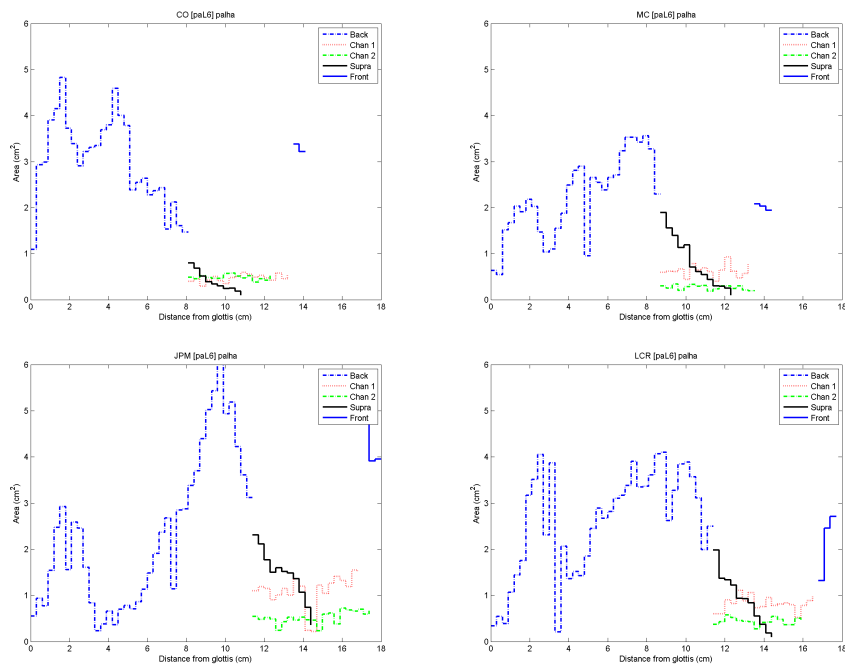


Figura 5.59: Funções de área do trato vocal para 2 falantes femininos (CO e MC) e dois masculinos (JPM e LCR) relativos ao /L/ produzido em [paL6].

Uma vez que estabelecemos a comparação entre a forma da língua em 3D para a lateral alveolar e palatal, será também interessante avaliar de que modo essa se reflete ao nível das funções de área. A figura 5.60 apresenta a sobreposição das funções de área para as duas consoantes, nas produções em posição intervocálica ([biliS e piL6]).

Note-se que algumas das diferenças que fomos apontando ao longo do texto se demonstram nestas imagens de forma bastante clara, nomeadamente a maior extensão do contacto para a palatal, que se estende até à zona alveolar. A diferença nas áreas, para as duas consoantes laterais, é também evidente. A posição mais elevada da língua na palatal conduz a áreas menores na cavidade oral para esta consoante (no [l] está baixa e recuada) e a anteriorização reflete-se em áreas maiores (para a palatal) a nível da faringe. A nível da glote as áreas são equivalentes para ambas as consoantes.

Tabela 5.19: Comprimento e área média, para as diferentes cavidades do [L], como produzido em [paL6] por quatro falantes; M (Masculino); F (Feminino).

Falante	Cavidade	Comprimento (cm)	Área (cm ²)	Desvio Padrão
CO (F)	Posterior	8.10	3.00	0.97
	Frontal	0.60	3.30	0.11
	Lateral1	5.40	0.47	0.08
	Lateral2	4.50	0.49	0.05
	Supra-lingual	3.00	0.38	0.22
MC (F)	Posterior	8.70	2.22	0.89
	Frontal	0.90	2.02	0.07
	Lateral1	4.80	0.63	0.13
	Lateral2	5.10	0.26	0.05
	Supra-lingual	3.90	0.80	0.57
LCR (M)	Posterior	11.40	2.59	1.15
	Frontal	0.90	2.16	0.74
	Lateral1	5.40	0.82	0.15
	Lateral2	4.80	0.42	0.13
	Supra-lingual	3.30	0.89	0.57
JPM (M)	Posterior	11.40	2.33	1.64
	Frontal	0.90	4.32	0.68
	Lateral1	5.70	1.11	0.35
	Lateral2	6.30	0.52	0.13
	Supra-lingual	3.30	1.44	0.56

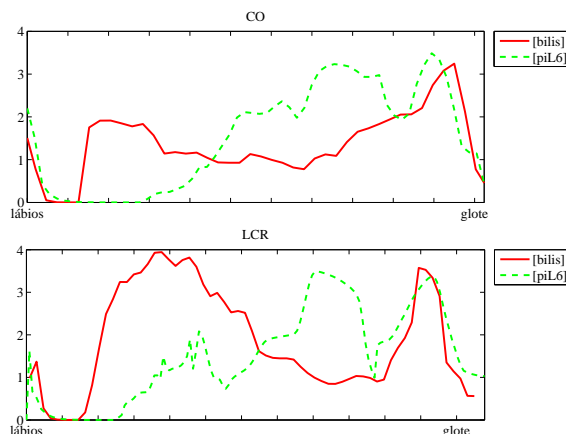


Figura 5.60: Sobreposição das funções de área do trato vocal na produção do [l] e [L] em posição intervocalica no contexto da vogal [i]. Exemplo para a falante CO e falante LCR.

5.5.2.4 Efeito do contexto vocálico

De forma similar ao que foi efetuado para a lateral alveolar, procurou-se avaliar, qualitativamente, o efeito da vogal na produção do /L/ (efeitos coarticulatórios), através da sobreposição dos contornos 2D da língua. Assim, começamos por sobrepor e comparar contornos, depois apresentam-se os IP (informação 2D) e finalizamos com as funções de área (3D). Na figura 5.61, pode-se observar a sobreposição dos contornos 2D, obtidos no plano médio-sagital, para seis dos falantes.

A análise das referidas imagens evidencia que a configuração articulatória da consoante

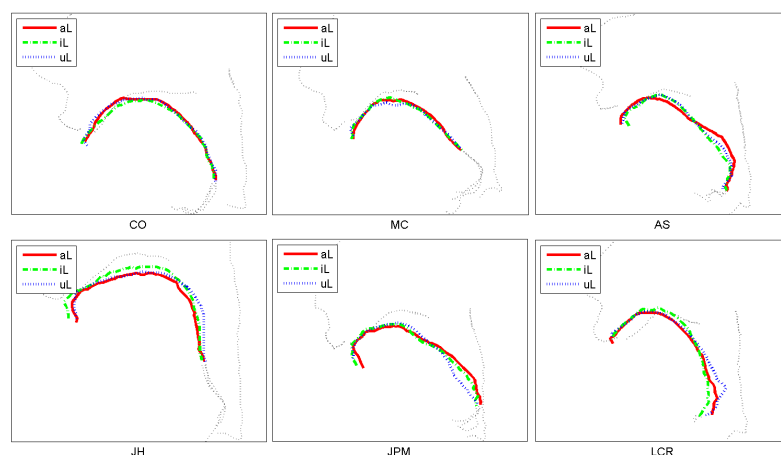


Figura 5.61: Efeito contextual em 2D - Sobreposição dos contornos 2D obtidos durante a produção do [L] no contexto das vogais cardinais para duas falantes femininas (CO e MC) e para os quatro falantes do género masculino (AS, JH, JPM e LCR).

parece ser pouco influenciada pela vogal que a antecede. Quando esse efeito é maior (como será o caso dos falantes JPM, AS e LCR) ocorre sempre atrás da área de contacto, sendo mais evidente, nestas imagens, a nível da raiz da língua, uma área da língua que não está envolvida na articulação. Nas falantes do género feminino (CO e MC) as diferenças entre os três contornos são subtis. Os IP's obtidos para todas as comparações de contornos e falantes são apresentados na tabela 5.20 e corroboram a informação obtida através da sobreposição de contornos permitindo agora uma quantificação dos efeitos coarticulatórios. De facto, são as falantes CO e MC que apresentam os IP's médios mais elevados, o que significa uma maior semelhança entre os contornos. O valor médio dos IP's, para todas os pares de comparações e todos os falantes, é de 0.730, valor que é semelhante ao IP médio obtido nas produções do /l/ quando se avaliou o efeito do contexto na mesma posição (ver secção 5.5.1.6). Também se pode observar que os contornos do /L/ em contexto de [a] e [u] são mais semelhantes entre si do que cada um deles com o [i].

Tabela 5.20: Valores dos IP calculados, em posição intervocálica para o /L/, para os pares de contornos [a-i], [a-u] e [i-u]

Género	Falante	a-i	a-u	i-u	Média
Fem.	CO	0.818	0.905	0.837	0.853
	MC	0.904	0.832	0.847	0.861
	ER	0.634	0.550	0.795	0.660
Mas.	AS	0.604	0.679	0.838	0.707
	JH	0.677	0.792	0.665	0.711
	JPM	0.726	0.587	0.692	0.668
	LCR	0.636	0.740	0.574	0.650
	Média	0.714	0.750	0.726	0.730

A avaliação dos efeitos contextuais em 3D, através da sobreposição das funções de área obtidas, coloca em evidência a existência de algumas diferenças entre a informação 2D e 3D. A

figura 5.62 inclui imagens de quatro falantes com sobreposição das funções de área para o /L/ no contexto das vogais [a, i, u]. Veja-se que, em algumas situações, as variações nas áreas são mais significativas que o revelado através da sobreposição de contornos em 2D. Para a falante MC, por exemplo, há uma variação nas áreas a nível da faringe que é impercetível nos dados 2D. A existência de *groove*, mais ou menos acentuado, na parte posterior da língua poderá conduzir a áreas muito diferentes e tal não é visível na informação a nível do plano médio-sagital. Por outro lado, no falante JPM, em que a informação em 2D indicava a existência de maiores diferenças, apresenta em 3D menos variação entre as três produções. Apesar destas diferenças, alguma da informação 2D é consubstanciada pelos dados 3D. É permitida menos variabilidade na zona do contacto e imediatamente atrás deste, porque serão estas as zonas que estarão mais constrangidas na produção do som, como tal, menos suscetíveis ao efeito da vogal adjacente.

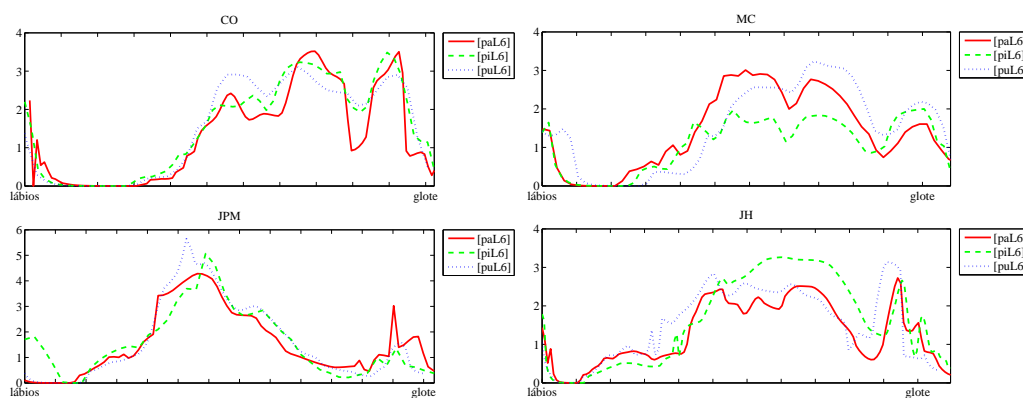


Figura 5.62: Funções de área do trato vocal para 2 falantes femininos (CO e MC) e dois masculinos (JPM e JH) relativos ao /L/ produzido no contexto das três vogais como em [paL6], [piL6] e [puL6].

5.5.2.5 Utilização da informação numa experiência de simulação para o /L/

Tendo como principal objetivo obter informação sobre a validade dos dados obtidos, em particular, no que diz respeito a algumas opções metodológicas adotadas (e.g., na determinação da área e extensão das passagens laterais), foi conduzida uma primeira experiência de simulação acústico-articulatória, na qual a autora esteve diretamente envolvida (segmentação e análise resultados). Os resultados serão apresentados de forma resumida, uma vez que não constitui um objetivo primordial deste trabalho. Para mais informação, poderá ser consultada a publicação Teixeira *et al.*, 2012b.

O *software* utilizado para as simulações aqui apresentadas é o *Vocal Tract Acoustic Response* - VTAR (Zhang *et al.*, 2003; Zhang e Espy-Wilson, 2004) que é de domínio público e corre em ambiente MATLAB. Este simulador contempla um modelo articulatório para o /l/, permitindo efetuar simulações com uma ou duas passagens laterais e cavidade supra-lingual separada. Por essa razão, achou-se adequado recorrer a este *software* para efetuar algumas experiências de simulação para o [L], uma vez que, como vimos, nesta consoante há também formação de passagens laterais e de cavidade supra-lingual. No entanto, caberá referir que, uma das limitações do VTAR é que não permite a simulação de passagens laterais assimétricas (em comprimento) como acontece, na

realidade, com a maior parte dos nossos falantes. Na literatura consultada, não existe qualquer referência a simulações com a lateral palatal. O modelo mais aproximado será para uma glide palatal, apresentado em Stevens (1998).

A realização desta experiência tornou necessário realizar novas segmentações do trato vocal para as produções dos quatro falantes que foram incluídos na experiência. Os cortes foram segmentados de 3 em 3 *mm* ao longo de todo o trato vocal e tornou-se necessário efetuar uma divisão do trato em várias cavidades (cavidades posterior, cavidade supra-lingual, passagens laterais direita e esquerda e cavidade frontal). A figura 5.63 apresenta um esquema, ilustrado com visualizações 3D, das diferentes cavidades consideradas para a simulação. A área ao longo das cavidades não é uniforme ao longo do seu comprimento (cf. as funções de área apresentadas na figura 5.59).

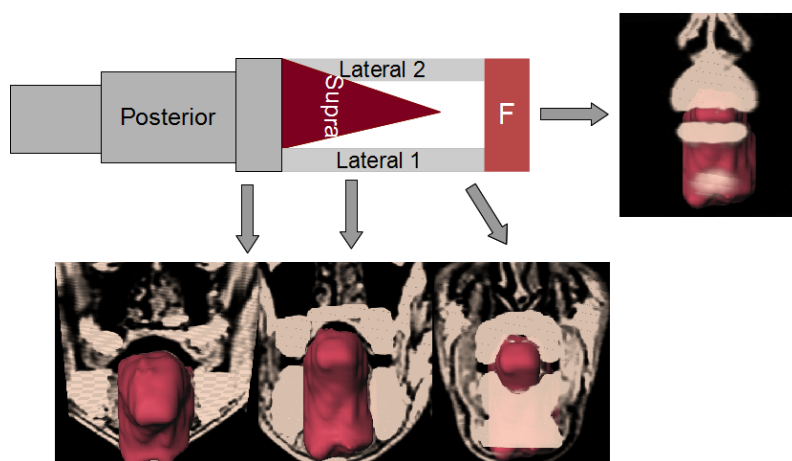


Figura 5.63: Representação em esquema das diferentes cavidades consideradas para a simulação e visualizações 3D da língua sobre um volume com as arcadas dentárias registadas que mostram as cavidades posterior, supra-lingual e frontal e as passagens laterais. Imagens do falante LCR durante a produção do [L] em palha.

Os CSO's, obtidos pelas segmentações ao longo do trato, correspondentes a cada uma destas cavidades tiveram que ser exportados de forma individualizada para ser possível efetuar as simulações no VTAR. Para além disso, tiveram que ser efetuadas algumas adaptações para que se pudesse efetuar a simulação com passagens laterais assimétricas, seguindo uma das abordagens proposta por Zhou (2009) para contornar problema semelhante (nesse caso para o /l/). As gravações de sinal acústico efetuadas em condições semelhantes às da RM (posição de supinação), para cada uma dos falantes, serviram para comparar o espectro acústico das produções naturais dos falantes e as simulações obtidas. Para este efeito, foram extraídos, de forma manual, os valores das três primeiras formantes e os zeros. Note-se que o processo de extração dos zeros a partir do espectro acústico é, de acordo com Zhou (2009) um processo pouco sistemático.

Uma vez que o VTAR tem sido utilizado para simulações com o /l/ e tendo ao nosso dispor dados para esta consoante, fez-se uma primeira experiência de simulação para esta consoante para um dos falantes (JPM). A função de resposta acústica (formantes e zeros) simulada foi muito

próxima da obtida para a produção natural do falante (cf. Teixeira *et al.*, 2012b) e apresenta um padrão formancial semelhante ao descrito para o Inglês Americano (Zhang *et al.*, 2003; Zhang e Espy-Wilson, 2004; Zhou, 2009).

Depois desta experiência, foram conduzidas as simulações para o /L/. Foi considerada a produção do /L/ no contexto da vogal [a]. A tabela 5.21 resume os resultados e apresenta os valores das formantes (F1, F2, F3) e zeros (em Hz), para o /L/, obtidos para as simulações conduzidas com o VTAR e para as produções naturais, produzidas pelos quatro falantes. Está também indicada a percentagem de erro entre as produções simuladas e as produções naturais dos falantes.

Tabela 5.21: *Comparação dos valores das Formantes (F1, F2, F3) e zeros (em Hz), para o [L], obtidos nas simulações efetuadas com o VTAR e as suas produções naturais, para quatro falantes.*

Falante	Simu/Nat	F1	F2	F3	Zero1	Zero2
CO	Simulado	468	2198	3135	3456	4033
	Natural	413	2442	3658	3165	4162
	Erro (%)	13	-10	-14	9	-3
MC	Simulado	446	1755	3102	3390	3623
	Natural	283	2229	3013	3024	4470
	Erro (%)	58	-21	3	12	-19
LCR	Simulado	440	1511	2564	3251	4055
	Natural	288	1842	3029	3453	4487
	Erro (%)	53	-18	-15	-6	-10
JPM	Simulado	463	1011	2653	3262	3451
	Natural	323	1739	2677	3132	4119
	Erro (%)	43	-42	-1	4	-16

Como se pode observar através da comparação entre os valores obtidos com as simulações no VTAR e as produções naturais dos falantes, as simulações permitem uma localização bastante precisa dos zeros introduzidos pelas passagens laterais e pela cavidade supra-lingual (erro absoluto médio de 7.8% para um dos zeros e de 12.0% para o outro), o que indicia que as áreas e extensão destas cavidades, obtidas a partir das imagens de RM, serão próximas das reais.

No que concerne a F2, foi no falante JPM que se verificou o maior desvio entre a simulação e o valor real (erro de 42.0%). Para os restantes falantes, o erro calculado para a segunda formante variou entre 10.0% e 21.0% sendo o erro absoluto médio para esta formante de 22.8%¹⁶.

Relativamente a F3, a percentagem de erro associada foi mais pequena (erro absoluto médio de 8.3%). O resultado que, no global, mais se afastou das produções naturais foi F1, com um erro absoluto médio de 41.8%. A única exceção, foi a falante CO, em que o erro foi de apenas 13.0%.

Para prever a localização dos zeros introduzidos pela presença das passagens laterais e pela cavidade supra-lingual, tornou-se necessário efetuar algumas simulações, em diferentes condi-

¹⁶Numa segunda tentativa de simulação efetuada para o falante JPM, ao invés de se considerarem os valores das áreas efetivamente medidas, considerou-se uma divisão da cavidade posterior em três secções, utilizando em cada uma delas a área média ao longo do seu comprimento. Com esta aproximação, o erro associado a F2 diminuiu para 12.0%, sendo equivalente ao obtido para os outros falantes.

ções. A figura 5.64 apresenta as funções de resposta acústica do trato vocal para dois dos falantes: um masculino (LCR) e outro feminino (CO).

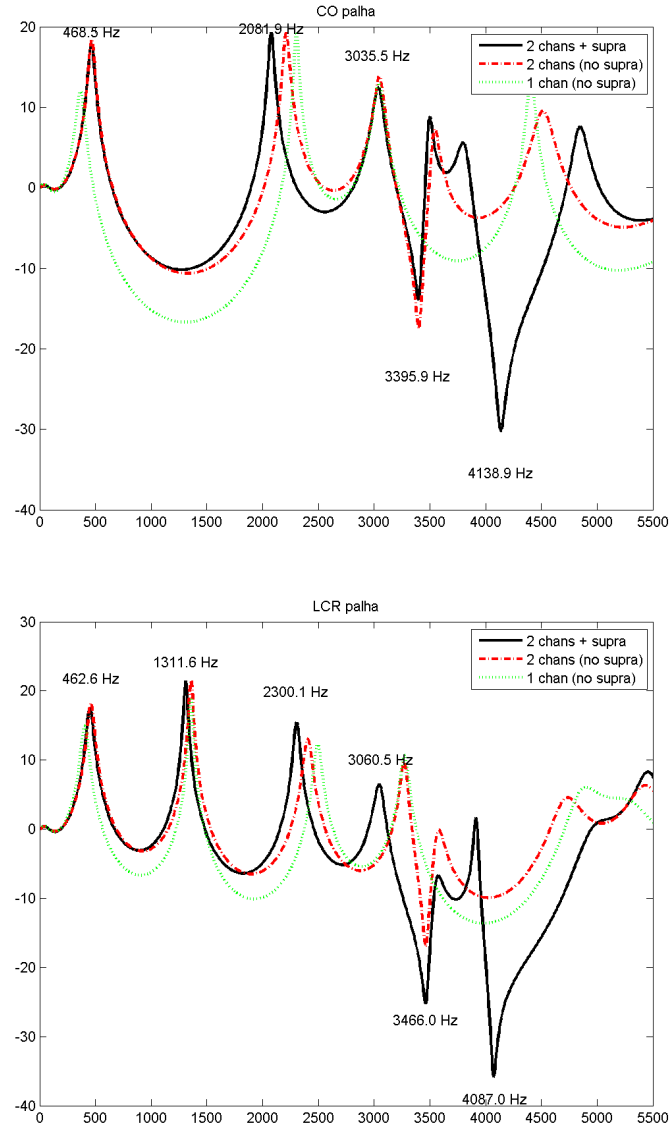


Figura 5.64: Função de resposta acústica do trato vocal obtida a partir de simulações efetuadas em diferentes condições: dois canais laterais e cavidade supra-lingual (preto), dois canais laterais sem cavidade supra-lingual (vermelho) e apenas um canal lateral sem cavidade supra-lingual (verde). Valores da ordenada em dB e abscissa em Hz .

Como se pode constatar, a simulação efetuada com a presença de apenas um canal lateral e ausência de cavidade supra-lingual não introduz nenhum zero no espectro (linha verde). A simulação com dois canais laterais e ausência de cavidade supra-lingual, faz surgir um zero na zona dos $3.0 - 3.5 kHz$ (linha vermelha). O outro cenário, mais semelhante ao que efetivamente se observa no nosso grupo de falantes na produção da palatal, inclui a simulação com presença de dois canais laterais e cavidade-supra lingual. Com estas condições, surge um segundo zero,

localizado um pouco acima dos $4kHz$. Estas simulações permitem-nos afiliar cada um dos zeros do espectro à cavidade que os origina: o primeiro é devido aos canais ou passagens laterais e o segundo é introduzido pela cavidade supra-lingual.

Os resultados obtidos são encorajadores, com excepção dos obtidos para F1, tendo em consideração as condicionantes do VTAR e o facto do modelo não incluir um módulo específico para a modelação da lateral palatal. Como já referimos, não temos forma de confrontar diretamente estes resultados de simulação para o /L/ com os de outros autores, por não conhecermos estudos similares efetuados para esta consoante. Todavia, uma vez que ambos os sons são laterais e partilham algumas características, poder-se-á estabelecer algum paralelismo com os resultados de simulação e com os modelos do trato vocal já existentes para o /l/, particularmente, no que diz respeito às dimensões das passagens laterais e da cavidade supra-lingual.

Assim, observamos que de forma semelhante ao reportado para a lateral alveolar (Prahler, 1998; Stevens, 1998; Zhang e Espy-Wilson, 2004; Zhou, 2009) também para a lateral palatal, a existência de um dos zeros no espectro acústico da consoante se ficará a dever à formação de canais laterais concorrendo para isso uma conjugação de fatores que, para além do comprimento total dos canais laterais, inclui a sua assimetria em extensão e em área. O outro zero é introduzido pela cavidade supra-lingual. Foi possível verificar que, pelo facto das áreas não serem uniformes ao longo da sua extensão e serem diferentes do lado direito e esquerdo e/ou assimétricas no seu comprimento, existe um zero entre F3-F5, mesmo com dimensões em comprimento (cf. tabela 5.19) inferiores às previstas por Prahler (1998) - para passagens com áreas uniformes o comprimento total das duas passagens teria que ser superior a 16 *cm*.

Também Zhang e Espy-Wilson (2004) e Zhou (2009) haviam já chegado à conclusão que era possível obter um zero entre os 3–5 *kHz*, com passagens laterais mais curtas, desde que estas fossem assimétricas e existisse estreitamento ou constrição alveolar. Zhou (2009) refere que para ser possível introduzir um zero abaixo dos 6 *kHz*, originado pela cavidade supra-lingual implica que esta tenha no mínimo 3 *cm* de comprimento. No nosso caso, relativamente à palatal, em todos os falantes a cavidade supra-lingual apresentou dimensões superiores a 3 *cm* e, em todas as situações, surgiu um zero nesta gama de frequências. Há um aspeto dos nossos resultados que difere do reportado por Zhou (2009) para a lateral alveolar. Este autor verificou que o zero introduzido pela cavidade supra-lingual ocorre a uma frequência mais baixa (2910 *Hz*) do que o zero atribuído às passagens laterais (4600 *Hz*).

Nas nossas simulações, observou-se precisamente o contrário para todos os falantes. Este facto pode dever-se a um comprimento e/ou áreas diferentes das passagens laterais no /l/ e no /L/, bem como a diferentes áreas da cavidade supra-lingual para as duas consoantes. É sabido que diferentes comprimentos e áreas podem deslocar consideravelmente a localização dos zeros (Zhang e Espy-Wilson, 2004; Zhou, 2009) no espectro. Por exemplo, Zhang e Espy-Wilson (2004) demonstraram que a não uniformidade da função de área para as passagens laterais produz um efeito semelhante ao aumento do seu comprimento, resultando numa frequência mais baixa para o zero.

No que diz respeito a F3, formante usualmente associada com às ressonâncias da cavidade frontal, anterior à zona da oclusão, o erro associado foi muito pequeno, o que de alguma forma valida as medidas obtidas a partir das imagens de RM (quer em área, quer em extensão) para esta cavidade. Note-se que a cavidade frontal para qualquer dos falantes foi bastante pequena (entre 0.6 a 0.9 *cm*) o que reforça a ideia de que a consoante é articulada muito anteriormente.

5.5.3 Comprimento do trato vocal

Não são abundantes na bibliografia dados relativos ao comprimento do trato vocal, em particular, para indivíduos do género feminino. Para falantes do PE, que seja do nosso conhecimento, não existe qualquer referência a estas medidas. Uma vez que, no nosso estudo, estes valores foram calculados, achamos oportuno reportá-los nesta secção de resultados. Assim, na tabela 5.22, apresentam-se os valores obtidos para três falantes do género feminino e para três do género masculino. Os valores apresentados referem-se ao valor médio obtido para o conjunto de nove produções, de cada falante, na produção do [l]. Uma vez que o comprimento do trato está de alguma forma relacionado com o biótipo do indivíduo (Story *et al.*, 1996; Fitch e Giedd, 1999) foi incluída, na tabela, a informação relativa ao peso e altura dos falantes.

Naturalmente, o comprimento do trato vocal é maior nos homens do que nas mulheres. Considerando a média dos três falantes masculinos (16.4 *cm*), verifica-se que o trato vocal destes é, em média, 3 *cm* maior do que nas mulheres (13.4 *cm*). Dentro de cada género, os valores foram bastante próximos, o que estará relacionado com uma constituição física semelhante.

O comprimento do trato vocal foi ligeiramente superior em produções do /l/ em posição de coda, em alguns dos falantes, mas tal não se verificou de forma sistemática (cf. tabela 5.23).

Tabela 5.22: *Comprimento médio do trato vocal, para seis dos falantes participantes no estudo, para as 9 produções, efetuadas para o /l/.*

Falante	Altura (<i>cm</i>)	Peso (<i>Kg</i>)	Média /l/ (<i>cm</i>)
CO (F)	160	50	13.3
MC (F)	160	63	13.6
ER (F)	150	50	13.5
JH (M)	183	75	16.3
LCR (M)	175	70	16.1
JPM (M)	190	98	16.9

5.6 Discussão

Na presente secção, procederemos à análise global e discussão dos resultados, tendo em mente os objetivos previamente delineados para esta parte do nosso trabalho e as questões que foram enunciadas na secção 6.3 relativamente às consoantes laterais. Esta análise contemplará, em primeiro lugar, uma reflexão relativa aos aspetos mais técnicos subjacentes a este trabalho (aquisição e processamento de imagem). Numa segunda parte, iremos debruçar-nos sobre as questões mais

Tabela 5.23: *Comprimento médio do trato vocal para seis falantes participantes no estudo, em função da posição silábica.*

Falante	início (<i>cm</i>)	medial (<i>cm</i>)	coda (<i>cm</i>)
CO (F)	13.1	13.4	13.2
MC (F)	13.6	13.5	13.5
ER (F)	13.5	13.4	13.5
JH (M)	16.3	16.3	16.2
LCR (M)	15.9	16.0	16.2
JPM (M)	17.0	17.0	16.9

direcionadas para o conhecimento fonético, interpretando e confrontando os nossos resultados com os dados já publicados sobre as consoantes laterais.

De forma genérica, podemos afirmar que os principais objetivos delineados para este trabalho foram atingidos. Foi elaborado e adquirido um *corpus* vocacionado para o estudo das consoantes laterais, que contemplou as principais posições silábicas em que as laterais podem ocorrer e foi considerado o efeito do contexto vocálico. O *corpus* foi adquirido num número significativo de falantes, incluindo falantes do género feminino e masculino, constituindo, muito provavelmente, uma das mais extensas bases de dados de RM existentes, a nível mundial, para o estudo das consoantes laterais.

Os protocolos otimizados, no âmbito do presente trabalho, para a aquisição das imagens de RM, em conjunto com a exploração de diversas técnicas de processamento de imagem revelaram-se adequados, possibilitando extrair informação de natureza muito variada e dados articulatórios (qualitativos e quantitativos) que viabilizaram a caracterização articulatória pormenorizada das consoantes laterais, um dos grandes objetivos desta investigação.

Algumas abordagens efetuadas no âmbito deste trabalho são totalmente inovadoras, como sejam a utilização de malhas poligonais para comparar sons em 3D ou a experiência de modelação acústico-articulatória, conduzida para a lateral palatal. As estratégias seguidas no contexto do presente trabalho são extensíveis ao estudo de outros sons e podem facilmente ser aplicadas a áreas de investigação mais clínicas.

Do ponto de vista técnico, foram ultrapassados alguns dos constrangimentos que se nos colocaram em estudos prévios (e.g., segmentação e co-registo das arcadas dentárias) e melhorados aspetos relacionados com a aquisição das imagens nestes estudos, profusamente apontados por outros autores, como sejam: reposicionamento do falante, tempo de aquisição muito longo, necessidade de efetuar várias repetições para adquirir imagens do trato vocal na sua totalidade e a falta de qualidade das reconstruções 3D, obtidas a partir de séries de cortes em 2D. Na realidade, a sequência otimizada para o estudo (3D-VIBE) permitiu-nos a aquisição de um volume de dados considerável em apenas 19s, o que constitui uma evolução considerável relativamente ao reportado em estudos semelhantes. Não houve necessidade de efetuar repetições do som para obter a configuração do trato vocal nem reposicionamento do falante. Obviamente que, nestas circunstâncias, a análise será mais complicada e existirão mais fatores de erro a considerar.

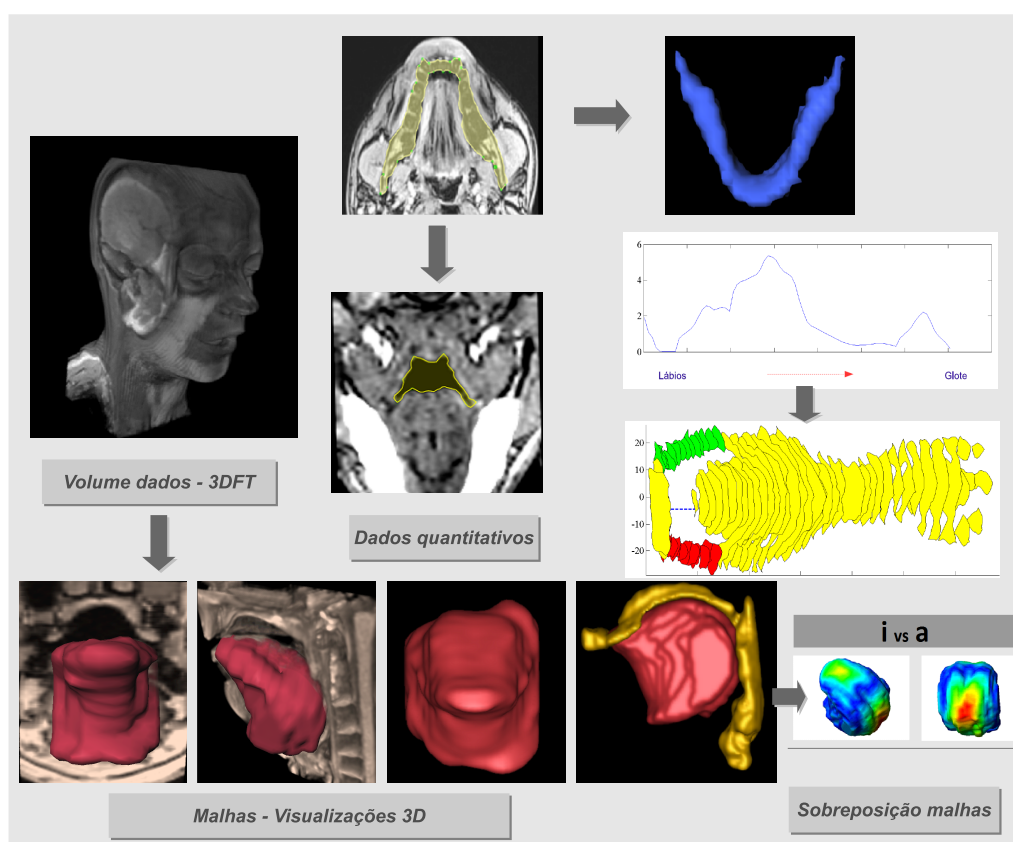


Figura 5.65: Tipo de informação retirado dos volumes de dados de RM

No que diz respeito às tarefas de processamento e segmentação de imagem foram testadas várias ferramentas e algoritmos, maioritariamente semi-automáticos, existindo atualmente algumas alternativas para a segmentação das diferentes estruturas (língua, véu palatino, trato vocal, arcadas dentárias), quer em imagens 2D quer em 3D. Neste domínio, afigurava-se como fundamental efetuar o co-registo das arcadas dentárias de cada um dos falantes para posterior modelação das passagens laterais formadas dos lados da língua. Ainda que não se tenham conseguido reconstruções perfeitas destas estruturas, foi possível ter uma ideia do espaço que ocupam na cavidade oral permitindo-nos orientar as segmentações e obter valores de comprimento e área que serão próximos das reais dimensões destas passagens.

No que diz respeito aos aspetos fonéticos, retomando agora as questões relativas às laterais que nos propusemos abordar e aprofundar neste trabalho, podemos constatar que, na articulação das duas consoantes laterais, se verifica a existência de aspetos comuns tal como atestado por Narayanan *et al.* (1997a, 1999) para outras línguas. As características fundamentais que definem esta classe de sons são as seguintes: 1) existência de compressão lateral da língua; 2) forma convexa do seu dorso para favorecer a passagem de ar lateralmente; 3) contacto da língua com a região alveolar (/l/) ou alveolo-palatal (/L/) a nível da linha média.

Existem, obviamente, também algumas diferenças que caberá notar: a extensão do contacto na linha média é, em geral, bastante maior para o /L/ do que para o /l/, sendo, em consequência, maiores (em comprimento) as passagens laterais no /L/ do que no /l/. Como a compressão da língua é mais marcada na palatal, as áreas das passagens laterais também são maiores para esta consoante. Na produção do /L/ existe sempre a formação de uma cavidade supra-lingual (separada) o que nem sempre se verifica para o /l/, particularmente para as falantes femininas. A palatal é produzida com avanço do corpo da língua, contrariamente à lateral alveolar, em que há recuo ou elevação da mesma.

Verificamos também que, estando garantidas as condições para a produção do /l/, é permitida alguma variação, entre os sujeitos. Genericamente, constatou-se a existência de uma maior variabilidade (áreas, forma da língua, dimensões das passagens laterais) nos falantes do género masculino do que nas falantes do género feminino. Na produção do /L/ a variabilidade encontrada entre géneros é menor.

Como já referido, para a caracterização da configuração gestual do /l/, interessa considerar, para além da existência de compressão lateral, o ponto, o tipo e a extensão da articulação, bem como a dimensão das passagens laterais.

Neste estudo, com um número considerável de falantes e para um *corpus* relativamente extenso, verificamos que as articulações laminais foram as preferidas para a maioria dos falantes e não se verificou nenhuma associação entre o tipo de articulação e a posição silábica. Estes resultados estão em linha com os obtidos por Narayanan *et al.* (1997a), que também não encontraram relação entre a zona da língua envolvida na constrição e a posição silábica e, consequentemente, com aspetos relacionados com a velarização da consoante. O ponto de articulação estará, como apontado por estes autores, mais associado com as características dos próprios falantes. Já Recasens e Espinosa (2005) referem que, na produção do /l/, os falantes têm preferência por uma

articulação apical em detrimento de uma articulação laminal, particularmente no /l/ velarizado, o que claramente não se verificou para os nossos falantes. Zhou (2009) refere que a articulação é apical no /l/ velarizado (em coda) e laminal no /l/ não-velarizado (em ataque). No entanto, caberá ressaltar que este último estudo inclui apenas um falante, o que não permite aferir a variabilidade que possa existir.

Relativamente à extensão do contacto alveolar, na linha média, apesar de se ter verificado uma tendência para este ser menor nas produções em coda comparativamente às efetuadas em ataque, essas diferenças não são tão marcadas como atestado por Narayanan *et al.* (1997a) e por Zhou (2009) para o Inglês americano. Estes autores reportam diferenças significativas na extensão do contacto alveolar entre os /l/'s produzidos em posição inicial e final, assinalando-as como muito maiores em ataque (*light-l*) do que em coda (*dark-l*).

A perda de contacto alveolar observou-se apenas numa das produções de um dos falantes (JH), em posição coda, não indiciando que seja uma tendência para o PE. Neste aspeto os nossos resultados estão de acordo com o reportado para o Inglês Americano (Narayanan *et al.*, 1997a) onde a perda de contacto alveolar não foi também um achado frequente.

Considerando as passagens laterais que se estabelecem de um ou dos dois lados da língua, verificamos que existem algumas diferenças, particularmente entre géneros, na dimensão das mesmas. Nas falantes femininas, as passagens são, em geral, mais curtas e apenas se formam quando se estabelece o contacto alveolar na linha média. Nos falantes masculinos, a variabilidade observada é maior sendo difícil identificar um padrão comum. Observou-se uma tendência geral para que as passagens laterais fossem maiores em posição de ataque do que em ataque intervocálico ou coda.

O efeito vocálico não parece introduzir diferenças dignas de registo na dimensão das passagens laterais. Há alguma assimetria entre as passagens direita e esquerda, mas esta deverá ser avaliada, individualmente, para cada falante.

Contudo, e ainda no que concerne às passagens laterais para a lateral alveolar, o panorama encontrado é complexo sendo difícil estabelecer, por exemplo, um relação entre a dimensão das passagens laterais e a posição silábica e suas consequências no espectro acústico.

Também não é fácil confrontar diretamente os nossos resultados com os obtidos por outros autores, para a lateral alveolar, pois na maioria dos outros estudos não foram consideradas tantas produções (contextos) ou falantes. Por exemplo, no estudo de Zhang e Espy-Wilson (2004) a informação diz respeito a um falante e a uma produção do /l/; o mesmo ocorre na investigação conduzida por Zhou (2009) em que entra apenas 1 falante com duas produções do /l/ (uma em posição de ataque e outra em posição de coda). No estudo conduzido por Narayanan *et al.* (1997a), são reportados dados relativos a quatro falantes, mas as passagens laterais só foram medidas (extensão e área) na zona do contacto alveolar.

Ainda assim, no que diz respeito às áreas das passagens laterais, os nossos valores estão de acordo com os obtidos por Narayanan *et al.*, que reporta áreas que oscilam entre um mínimo de 0.06 cm^2 e um máximo de 0.43 cm^2 . Na experiência de modelação de Zhang e Espy-Wilson (2004), as áreas médias (de cada uma das passagens laterais) consideradas para simulação (0.90

cm^2 e $1.0\ cm^2$) são bastante acima das áreas médias por nós obtidas (área média máxima foi de $0.36\ cm^2$). Todavia, esta diferença pode ser justificada pela ausência de co-registo das arcadas dentárias do falante no estudo supracitado, o que, de acordo com os próprios autores, resulta em áreas sobreestimadas.

Relativamente ao comprimento das passagens laterais, no estudo de Zhang e Espy-Wilson (2004) a dimensão das passagens laterais obtida a partir das imagens de RM foi de $4.2\ cm$. No entanto, para efeitos de simulação, os autores consideraram uma passagem lateral com $5.0\ cm$ e outra com $3.4\ cm$ para averiguar o impacto da assimetria no comprimento no espectro acústico. No nosso estudo (cf. tabela 5.11), com exceção de duas produções do falante JPM (em [sal6] e [lit4u]), as passagens laterais apresentaram comprimento bastante menor (variando entre 0.2 e $3.2\ cm$).

Uma das hipóteses que levantamos relativamente à variabilidade (articulatória) observada nos nossos resultados no que diz respeito às passagens laterais é que esta poderá não ter impacto significativo no espectro acústico da lateral alveolar, particularmente, na distinção ataque/coda. Ainda que com comprimento diferente (assimetria), as passagens laterais serão demasiado pequenas para provocar a introdução de zeros no espectro, na gama de frequências entre F3-F5. Desta forma, uma diferença articulatória poderá resultar num *output* acústico semelhante. Também, as diferenças de dimensão no comprimento das passagens laterais entre os falantes femininos e masculinos, deverão ser analisadas e interpretadas tendo em consideração que o comprimento total do trato vocal, entre géneros, é diferente. Assim, para que se possa ir para lá de especulações, revela-se necessário retomar este assunto pensando em formas de efetuar normalização e, conduzir mais algumas experiências de modelação acústica, semelhantes às já efetuadas para a palatal.

Um aspeto que merece também a nossa atenção está relacionado com a velarização do /l/ e com a existência, ou não, de alofones posicionais, para o /l/, que como já referimos é ainda um assunto que divide os foneticistas que se dedicam ao estudo do PE.

Tendo em consideração o efeito da posição silábica, baseados na informação 2D (IP calculados para os diferentes contornos) e nas funções de área obtidas nas diferentes posições, podemos verificar que a configuração articulatória do /l/ produzido em ataque e coda não será exatamente igual mas as diferenças encontradas entre estas são subtis, o que apontará para a hipótese de existência de apenas um /l/, independentemente da posição silábica. Tendo em atenção as suas características articulatórias será possível afirmar que o /l/ no PE é velarizado ou faringealizado. De facto, em qualquer das posições consideradas, o dorso da língua assume sempre uma posição bastante recuada dentro da cavidade oral originando uma segunda constrição que pode ocorrer na zona velar/uvular ou, na zona da faringe. Esta segunda zona de constrição dará origem a duas cavidades, atrás da constrição principal, tal como apontado por Recasens (2012) para o /l/ velarizado.

Não se observou qualquer associação entre apicalidade e velarização, tal como defendido por Recasens e Espinosa (2005), Recasens (2012) e por Zhou (2009). Verificou-se, contudo, que as articulações, independentemente da posição ou contexto, foram sempre muito anteriores, ocorrendo na zona dental e na zona de transição dento-alveolar mais do que na zona alveolar propriamente

dita, o que favorece a velarização do /l/ (cf. Recasens, 2012).

A extensão do contacto na linha média, com excepção de poucas produções é semelhante à reportada por Narayanan *et al.* (1997a); Zhou (2009) para o /l/ velarizado. No caso do *light-l*, para o Inglês Americano, a extensão do contactos na linha média é significativamente maior.

No respeitante aos efeitos coarticulatórios, a nível da língua, verifica-se que o /l/ não é imune a efeitos coarticulatórios como são algumas consoantes fricativas /S, Z/ (veja-se Teixeira *et al.*, 2011) mas não será tão permeável a efeito contextual como as consoantes oclusivas (e.g., [p] ou [b]). Há, na realidade, alguma influência do contexto vocálico sobre a consoante. No entanto, o facto de os efeitos coarticulatórios serem equivalentes nas várias posições é mais um fator a favor da existência de apenas uma configuração articulatória para o /l/ e que será tendencialmente velarizada (Recasens, 2012).

Assim, globalmente, os resultados articulatórios obtidos vão ao encontro das descrições de Stevens (1954), que refere para o PE uma variedade fortemente velarizada do /l/ e, também, das afirmações mais recentes de Emiliano (2009). Também corroboram os dados acústicos obtidos por Andrade (1999), para falantes da zona de Lisboa, que indiciam a presença de um /l/ velarizado não só em posição final, mas também em posição inicial. Estão também em linha com resultados acústicos de dois estudos acústicos, mais recentes, conduzidos por Marques (2010) e Monteiro (2012) que incluem, tal como o presente estudo, falantes da região norte e centro do país. Desta forma, os dados apontam para a hipótese da velarização do /l/, extensiva a todo o território nacional, como já havia sido sugerido por Oliveira (2009).

No tocante ao /L/ não havia, até ao momento, dados que permitissem uma descrição articulatória detalhada desta consoante e neste aspeto os resultados deste trabalho constituem uma novidade no que diz respeito ao PE. Em relação aos gestos articulatórios para a palatal verificamos que a articulação da palatal é bastante mais anterior do que o apontado nas descrições tradicionais. A extensão da articulação é bastante extensa o que em conjunto com a compressão lateral da língua condicionará passagens laterais maiores (comprimento e área) que as observadas para a lateral alveolar.

A comparação do ponto de articulação desta consoante com o da nasal [J] leva-nos a sugerir que o /L/ deva ser incluído na categoria de consoantes alveolo-palatais e não na das palatais propriamente ditas. De facto, se o /J/ no PE é articulado mais atrás do que o /L/ (cf. Martins, 2007 e Oliveira, 2009) e é considerada uma consoante alveolo-palatal não fará sentido que o /L/ seja classificada como palatal. Estes dados confirmam, os resultados prévios para o PE obtidos num menor número de falantes por Martins (2007) e Oliveira (2009) e estão de acordo com o que tem vindo a ser defendido (Recasens, 1990; Recasens *et al.*, 1993; Recasens e Espinosa, 2006) para outras línguas Românicas em que esta consoante ainda ocorre (e.g., Catalão de Maiorca) e com as descrições encontradas para o PB (Cagliari, 1977b; Reis e Espesser, 2006).

Ao contrário do que se está a verificar para algumas línguas Românicas (cf. Colantoni, 2004) e também referido para alguns dialetos do PB não há qualquer indício que, no PE, esta consoante se esteja a converter numa glide palatal ou a sofrer um processo de iotização ou de apagamento (Stein, 2011). No entanto, remetemos estas questões para foneticistas e linguistas.

Finalmente, apenas uma referência aos valores encontrados para o comprimento do trato vocal. Os valores estimados para o comprimento do trato vocal obtidos no nosso estudo (16.4 *cm* para os homens e de 13.4 *cm*) estão de acordo com o padrão (maior nos homens do que nas mulheres) e com os valores usualmente reportados na literatura (Goldstein, 1980; Stevens, 1998; Fitch e Giedd, 1999). Goldstein (1980) e Stevens (1998) apresentam um valor estimado de 16.9 *cm* para o comprimento do trato vocal em indivíduos do género masculino e 14.1 *cm* para indivíduos do género feminino (em repouso). Os autores ressaltam que o comprimento do trato pode variar, para os diferentes fonemas, em função da posição dos lábios (arredondamento ou não) e da posição da laringe. Comparativamente aos dados reportados por Narayanan *et al.* (1997a), obtidos com RM para falantes americanos dos dois géneros, verifica-se que, em particular, no caso das falantes do género feminino os valores por nós obtidos são bem inferiores aos calculados no estudo supracitado (17.2 *cm* e 15.5 *cm*) o que deverá muito provavelmente¹⁷ refletir as diferenças de biótipo entre as nossas falantes e as participantes nesse estudo.

Neste capítulo congregou-se toda a informação relativa às etapas de aquisição, processamento, extração de dados e apresentação dos resultados obtidos com esta parte do nosso estudo e que permitiram efetuar uma descrição articulatória das consoantes laterais ancorada em dados qualitativos e quantitativos. Como ficou patente, o estudo destes sons é bastante complexo com inúmeras vertentes que têm que ser tidas em consideração para a sua análise. Ainda que o presente trabalho possa ter dado algum contributo para melhor conhecer a articulação destes sons haverá ainda alguns aspetos que deverão continuar a ser trabalhados e melhorados.

¹⁷Esta é uma suposição nossa uma vez que no estudo não é referido a altura nem o peso dos falantes.

Capítulo 6

Contributos da RM em Tempo Real para o Estudo das Vogais Nasais do Português

One longstanding challenge in speech production research has been the ability to examine in real-time the changes in the shaping of the vocal tract...

NARAYANAN, S (2004)

O presente capítulo, congrega toda a informação relacionada com a obtenção de imagens de RM em tempo real para o estudo das vogais nasais do PE, visando contribuir para um melhor conhecimento das suas características dinâmicas. Assim, após um breve enquadramento relacionado com aquisição de imagens de RM em tempo real (secção 6.1), segue-se a apresentação de alguns estudos com RM que, tal como o nosso, se focalizam nos aspetos dinâmicos da nasalidade vocálica (secção 6.2). Apresentam-se depois os objetivos subjacentes a esta parte do nosso trabalho, elencando também as questões que nos propomos abordar (secção 6.3). De seguida, na secção 6.4, damos conta de toda a metodologia seguida, com a descrição do *corpus*, dos falantes, das etapas diretamente relacionadas com a aquisição de imagem e o resumo das abordagens de processamento efetuadas, que conduziram à obtenção de informação qualitativa e quantitativa. Posteriormente, apresentam-se os resultados já alcançados (secção 6.5). Na última parte do capítulo (secção 6.6), discutem-se os resultados, enquadrando-os com dados anteriormente coligidos para o PE, obtidos através de outras técnicas e, sempre que se justifique, com o conhecimento existente para outras línguas.

6.1 RM em tempo real

No capítulo 2, foi apresentada uma visão global da evolução tecnológica a que se assistiu na área da RM, nos últimos anos, sem que se tenha especificado qualquer técnica em particular. Nesta breve introdução ao capítulo 6, mencionaremos alguns aspetos diretamente relacionados com a

aquisição de imagem em tempo real, fazendo alusão a terminologia específica e referindo as suas principais aplicações.

Durante alguns anos, a aquisição de imagens em tempo-real (ou quase em tempo-real) com RM era uma utopia. A demonstração que era viável adquirir imagens dinâmicas, em tempo real, foi apresentada por Riederer *et al.* (1988), utilizando uma sequência FLASH modificada, que permitiu a aquisição de uma imagem completa num segundo. A sua aplicação na prática diária é bastante mais recente fruto das melhorias a nível da resolução temporal permitindo, em consequência, alargar o seu leque de aplicações.

Na atualidade, a aquisição de imagens de RM em tempo real com razões de amostragem consideravelmente elevadas (mais de 30 *frames/s*) constitui uma realidade (veja-se, apenas como exemplo, Uecker *et al.*, 2010b). Quando se pretende um número *frames* elevado, a aquisição é efetuada em modo 2D e, geralmente, limita-se a um único corte. Estabelecendo compromissos, com degradação da resolução temporal, é possível obter uma maior cobertura anatômica com maior número de cortes (Bernstein *et al.*, 2004; Kim *et al.*, 2012). Ainda que com limitações a nível da qualidade de imagem, é viável efetuar aquisições 3D em tempo real (Zhu *et al.*, 2013b). A aquisição de imagens de forma ainda mais rápida e a obtenção de maior cobertura anatômica é ambicionada e defendida por muitos autores, sendo certo que continuará a constituir um grande desafio do ponto de vista tecnológico, condicionado, em grande medida, pelos limiares fisiológicos (e.g., SAR e ENP).

Será pertinente, antes de avançar, esclarecer alguns aspetos relativos à terminologia usualmente utilizada nas aquisições de imagem em tempo real, nomeadamente no que diz respeito à resolução temporal e ao próprio significado do que é imagem em “tempo real”, uma vez que a sua utilização por vezes pouco precisa pode induzir algumas dúvidas. O termo “tempo real” pode ser alvo de diferentes interpretações considerando os diversos sistemas de imagem. Usualmente, é especificado em função do número de imagens ou *frames* que podem ser obtidas num segundo. No entanto, relativamente ao que é considerado o número de *frames* necessário para ter uma “verdadeira” aquisição em tempo real, muitos autores consideram que, ao invés de se definir um número fixo de imagens a alcançar, se deve assumir que uma imagem é, ou tem informação em tempo real, se o processo de aquisição de dados e formação da imagem coincidir com a escala temporal do fenómeno que se quer estudar (Bernstein *et al.*, 2004; Nayak e Hu, 2005; Riederer *et al.*, 2001)¹.

Assim, a obtenção de 1 ou 2 imagens/s pode ser considerada em tempo real e adequada para o estudo dinâmico de uma articulação (e.g., joelho), monitorizar a chegada de um *bolus* de meio de contraste para posterior planeamento do exame (vulgarmente, em estudos angiográficos), observar os movimentos do diafragma durante a respiração ou guiar a progressão de uma agulha de biópsia num procedimento de intervenção mamária. No entanto, noutras situações, particularmente, no âmbito de estudos cardíacos (e.g., estudo da função ventricular ou cateterização cardíaca), os requisitos para a frequência de amostragem são muito mais exigentes sendo necessário obter informação a cada 50-60 *ms* o que corresponderá, no mínimo, à obtenção de 17-18 imagens/s.

¹No entanto, é frequente a referência de “tempo real” como sendo de 24-25 *frames/s* que é a frequência de amostragem utilizada em filmes e vídeos.

No que diz respeito à resolução temporal, Bernstein *et al.* (2004) consideram que deve haver uma distinção entre o que designam por resolução temporal verdadeira (medida em *frames/s* ou *Hz*) e a resolução temporal efetiva ou aparente. Segundo a mesma fonte, para uma aquisição 2D, o cálculo da resolução temporal verdadeira é efetuado recorrendo à seguinte expressão:

$$R = \frac{1}{N_{Shot} \times TR \times NEX} \quad (6.1)$$

onde N_{Shot} é o número de pulsos de RF (intervalos de TR) necessários para preencher todo o espaço K e NEX é o número de excitações. Numa aquisição 2D convencional, o N_{Shot} é igual ao número de passos de codificação de fase. Todavia, o termo N_{Shot} é mais abrangente uma vez que contempla o cálculo da resolução temporal quando se utiliza modo de aquisição EPI, espiral, radial ou TSE. Por exemplo, para uma técnica *Single-Shot* (SS) o $N_{Shot} = 1$. A utilização de FOV rectangular, imagem paralela ou técnicas de preenchimento parcial do espaço K podem conduzir a um aumento da resolução temporal verdadeira, uma vez que possibilitam a redução do número de passos de codificação de fase necessários ao preenchimento da matriz. A resolução temporal aparente ou efetiva representa o número de imagens reconstruídas por segundo ou *frame rate*. Esta pode ser superior à resolução temporal verdadeira se forem utilizadas técnicas de reconstrução como *keyhole* ou *sliding windows*. Note-se, no entanto, que a resolução temporal verdadeira não é alterada (Bernstein *et al.*, 2004; Wieben, 2007).

Do ponto de vista dos requisitos tecnológicos, tendo em atenção o exposto por Riederer *et al.* (2001), um sistema para aquisição de imagem em tempo real terá que: i) permitir a aquisição de imagem de forma contínua²; ii) reconstruir e apresentar as imagens de forma rápida e, iii) ter possibilidade de alterar parâmetros de aquisição de dados em função das imagens reconstruídas. Relativamente a este último aspeto, é muito importante, particularmente se for necessário ajustar parâmetros como o FOV ou plano de aquisição, durante a aquisição, ter em atenção o tempo de latência total (*lag* ou *latency*). Esta define-se como o intervalo de tempo que decorre entre o acontecimento e a visualização da imagem, que deverá ser o mais curto possível (usualmente entre 0.1 a 0.5 s). Um dos fatores que concorre para o cálculo da latência é o tempo de reconstrução da imagem, que não é desprezível, mesmo em sistemas mais avançados. Se, por um lado o poder computacional é maior, por outro a utilização de matrizes mais elevadas, imagem paralela, maior número de canais recetores podem contribuir para o aumento do tempo de reconstrução das imagens (Bernstein *et al.*, 2004).

Considerando a utilização da RM em tempo real, em contexto clínico, vários autores apontam-na como preponderante no âmbito de estudos cardíacos, abrindo a oportunidade de avaliar doentes com arritmia, uma vez que não é necessária realização de aquisição sincronizada com ECG (Earls *et al.*, 2002; Nayak e Hu, 2005; Uecker *et al.*, 2010b; Zhang *et al.*, 2010b). Para além disso, apresenta grande potencial para melhorar a qualidade destes estudos na avaliação de isquémia do miocárdio, na doença pericárdica e valvular, no estudo da função ventricular (Kaji *et al.*, 2001; Earls *et al.*, 2002) e na avaliação de patologia cardíaca congénita. Nas patologias congénitas,

²Por existirem algumas semelhanças com os sistemas fluoroscópicos convencionais, há quem utilize a designação de Fluoroscopia por RM.

a RM em tempo real será particularmente benéfica na avaliação de doentes pediátricos permitindo, por exemplo, reduzir os tempos de anestesia necessários à realização dos estudos (Nayak e Hu, 2005). Há também grande interesse na sua utilização para monitorização de procedimentos de intervenção (e.g., biópsias) ou no decurso de cirurgias.

Outras áreas em que a RM em tempo real tem merecido considerável atenção nos últimos anos, a julgar pelo número de publicações que têm vindo a surgir, são: 1) a avaliação dos mecanismos de deglutição e mastigação (Anagnostara *et al.*, 2001; Barkhausen *et al.*, 2002; Breyer *et al.*, 2009; Zhang *et al.*, 2012); 2) avaliação do impacto de intervenções cirúrgicas na produção de fala, por exemplo, em indivíduos glossetomizados (Mády *et al.*, 2002; Zhou *et al.*, 2013); 3) avaliação de indivíduos com apneia do sono (Kim *et al.*, 2013); 4) monitorização de pacientes com fendas palatinas e incompetência velo-faríngea (Drissi *et al.*, 2011; Bae *et al.*, 2011; Perry *et al.*, 2013). Sell e Pereira (2012) afirmam mesmo que a RM em tempo real poderá vir a ter um papel fulcral na avaliação das patologias relacionadas com o véu palatino e mecanismo velo-faríngeo:

“The gold standard in the current investigations of multiview videofluoroscopy and nasoendoscopy have particularly focus on velar elevation and LPW movements. It is possible that real time MRI will revolutionize this, leading to a re-conceptualization of whole of this area of diagnostics and a completely different focus within investigations”
(Sell & Pereira, 2012, p.162)

Apesar das limitações da RM na visualização de estruturas calcificadas, particularmente dos dentes, tem sido crescente a sua aplicação na área da ortodontia. Veja-se, por exemplo, os estudos de Karacay *et al.* (2006) e de Görgülü *et al.* (2011) que utilizam RM em tempo real para avaliar o impacto da correção cirúrgica na postura da língua e nos movimentos de deglutição, em pacientes com mordida aberta e em casos de má oclusão.

No âmbito da produção e modelação de fala, como já referido, a possibilidade de obter imagens em tempo real representa uma mais valia inquestionável, uma vez que a fala é, inerentemente, um processo dinâmico e a modelação do trato vocal assente nessas características, será com toda a certeza, mais fidedigna (Demolin *et al.*, 1997; Engwall *et al.*, 2006; Narayanan *et al.*, 2004; Bresch *et al.*, 2008a; Kim *et al.*, 2010; Proctor *et al.*, 2013).

6.2 RM tempo real aplicada ao estudo da nasalidade vocálica

Apesar da quantidade de estudos com RM em tempo real na área da fala ser já relevante, escassas foram as referências encontradas que utilizam aquisições com RM em tempo real para estudo das vogais nasais³.

³Tendo em atenção o descrito nos capítulos 2 e 3, a maior parte dos estudos envolvendo aquisições com RM em tempo real, na área da produção de fala, está a ser conduzido pelo grupo SPAN da Universidade da Califórnia e visam essencialmente a língua Inglesa. Nesta língua existem consoantes nasais mas não existem vogais nasais que contrastem fonologicamente com as vogais orais (como no português ou francês) ocorrendo nasalização das vogais por coarticulação.

Sem grandes surpresas, verificamos que os estudos existentes são para o Francês (Delvaux *et al.*, 2002a; Delvaux, 2003; Proctor *et al.*, 2013) e para o PB (Gregio, 2006; Medeiros e Demolin, 2006). Cabe realçar que o estudo de Delvaux *et al.* (2002a) inclui, para além dos dados articulatórios obtidos com RM, dados aerodinâmicos e acústicos. Para o PB, as investigações com utilização de RM não são abundantes, no entanto existe um trabalho vasto que inclui estudos acústicos (e.g., Souza, 1994 ou Campos, 2009), aerodinâmicos (Medeiros, 2011) ou de fibroscopia (Lovatto *et al.*, 2007) que abordam diversas questões relacionadas com a produção das vogais nasais.

No estudos de Delvaux (Delvaux *et al.*, 2002a; Delvaux, 2003), o *corpus* RM em tempo real foi adquirido em quatro falantes, dois do género feminino e dois do género masculino, todos falantes nativos do Francês (Bélgica). O *corpus* contemplou 28 pseudo-palavras, contendo as quatro vogais nasais e as correspondentes orais, em diferentes contextos fonológicos (oral ou nasal). Para estabelecer a comparação, através da sobreposição de contornos, entre vogais nasais e orais, foram considerados os seguintes contextos:

- p_# - vogais orais e nasais precedidas da oclusiva /p/;
- NVN *vs* NV - vogais orais fortemente nasalizadas por coarticulação comparadas com vogais nasais, em contexto nasal;
- VVV - todas as vogais orais, nasalizadas e nasais, independentemente da sua posição.

O protocolo de aquisição foi o desenvolvido e apresentado em Demolin *et al.* (2000). A gravação do sinal acústico foi efetuada com recurso a um microfone ótico (MO). Os dados acústicos foram utilizadas apenas para alinhar e associar cada imagem de RM a 200 *ms* de sinal, que corresponde à resolução temporal conseguida. Cada segmento foi repetido três vezes, com aquisição de imagem no plano médio sagital durante um período de aquisição de 15 *s*, à razão de 5 *frames/s*. Foi extraído, através de traçado manual, um contorno relativo a uma imagem (parte média) de cada repetição. Todos os contornos referentes a uma determinada vogal, para um determinado falante, foram sobrepostos, permitindo estabelecer comparação entre vogais orais, vogais nasais e nasalizadas. A análise efetuada permitiu, segundo os autores, delinear as seguintes conclusões:

- O contraste fonológico entre vogais nasais e orais, no Francês, não resulta só de um abaixamento do véu palatino, mas também de uma série de ajustes articulatórios (e.g., arredondamento labial, recuo e abaixamento da língua, grau de abertura do véu palatino) que variam com as diferentes vogais;
- A posição do véu palatino sofre alterações com as diferentes vogais nasais, com os falantes e nos diferentes contextos estudados;
- As vogais orais são ligeiramente nasalizadas em contexto _N, fortemente nasalizadas em contexto N_, particularmente as vogais fechadas, e completamente nasalizadas em contexto N_N;

- Em contexto p__, a posição do véu palatino varia com o grau de abertura da vogal oral, mais alto para as vogais fechadas e mais baixo para a vogal aberta.

O trabalho de Proctor *et al.* (2013) visou estudar os padrões de coordenação do véu palatino e da língua na produção de consoantes e vogais nasais do Francês. Para comparação, foram incluídas as consoantes nasais, produzidas por falantes do inglês, em contextos similares. O *corpus* foi adquirido em dois falantes do Francês (um masculino e outro feminino) e em dois falantes do Inglês, do gênero masculino (um americano e outro australiano). Incluiu as vogais nasais do francês e consoantes nasais das duas línguas, em posição medial e final de frase, no contexto de oclusivas e fricativas.

Foi efetuado um corte no plano médio sagital com resolução de 68×68 *pixels* e *frame rate* de 33 *frames/s*. O protocolo RM seguido é o apresentado em Narayanan *et al.* (2004) com pequenas alterações. As imagens foram alinhadas com o sinal áudio (obtido com MO) para permitir a marcação de quatro referências, em sílabas de interesse, que permitiram extrair seis parâmetros temporais: “(i) *onset to vocalic target (V)*; (ii) *onset to initiation of velic lowering (Vel)*, (iii) *onset to tongue tip closure (TT)*; (iv) *vocalic target to initiation of velic lowering*; (v) *vocalic target to tongue tip closure*; and (vi) *initiation of velic lowering to tongue tip closure*.” (Proctor *et al.*, 2013 p. 578).

Os autores verificaram que os padrões de coordenação entre o véu palatino e os articuladores orais são diferentes nas vogais e consoantes nasais do francês e que o padrão de coordenação para as consoantes nasais do inglês é semelhante ao das vogais nasais do francês. Reafirmam a importância da RM em tempo real no estudo do véu palatino e dos padrões de coordenação temporal nos sons nasais.

Para o PB, o estudo conduzido por Gregio (2006), incluiu um falante do gênero feminino de 22 anos, natural de S. Salvador da Baía e residente em S. Paulo desde os 18 anos. Do *corpus* fizeram parte as vogais orais e nasais do PB produzidas de forma isolada e sustentada.

As aquisições foram efetuadas num equipamento GE de 1.5T, aquisição no plano sagital, utilizando uma espessura de corte de 8 *mm*. Foram obtidas quatro séries do trato vocal supra glótico, com a duração de 17 s, dando origem a 100 imagens por série, sendo referida uma resolução temporal de 5.8 *frames/s*. O sinal acústico foi gravado com um microfone unidirecional normal. Foram efetuadas várias tentativas de filtragem do ruído com sucesso relativo. Os resultados foram obtidos através da análise de vídeos efetuados para o efeito. Algumas imagens foram impressas e, nessas, foi realizado um traçado manual das estruturas anatómicas, permitindo observar e comparar a posição do dorso da língua, do véu palatino, dos lábios e da mandíbula. Para além disso, são apresentados, para as diferentes vogais, conjuntos sequenciais de imagens no plano sagital.

Como resultados, são referidas diferenças entre as vogais orais e nasais. No que concerne à postura do dorso da língua, a autora verificou que o dorso está mais anteriorizado na produção do [6 ~] do que na produção da correspondente oral. Na produção de [u ~] e [o ~], o dorso está mais retraído do que na oral correspondente, ainda assim mais pronunciado no [u ~], com

repercussões a nível da cavidade faríngea.

Para além disso, nas vogais nasais observaram-se posturas diferentes dos articuladores no decorrer da sua produção, sendo identificadas três fases distintas: uma fase oral, uma fase nasal e uma fase nasal com movimento do dorso língua.

Na fase oral, o véu palatino está elevado ou ligeiramente abaixado. Esta fase oral é comum a todas as nasais, mas a sua posição varia com a vogal nasal devido à necessidade de ajustes entre o véu palatino e a língua para obtenção da abertura velo-faríngea. Na segunda fase, a fase nasal, o posicionamento do dorso da língua mantém-se e há abertura da passagem velo-faríngea para permitir o acoplamento da cavidade nasal. Na terceira fase, existe um movimento do dorso, particularmente da região anterior, em direção à região palato-alveolar. Segundo Gregio (2006), esta terceira fase observada na produção das vogais nasais poderia corresponder ao murmúrio nasal.

É referido, como um dado relevante para o PB, que na produção da vogal nasal, o véu palatino começa por efetuar um movimento de fecho (fase oral) e não de abertura como seria de esperar:

“...este achado parece contrariar o argumento de ser a primeira fase um movimento para abrir a cavidade nasal, pois, inicialmente o palato mole fecha a cavidade nasal (primeira fase) e, posteriormente a abre (segunda fase)...” (Gregio (2006), p.66)

Gregio (2006) refere ainda que os resultados do estudo indiciam a não existência de vogais puramente nasais, como acontece no Francês, devido à existência de modificações na postura dos articuladores durante a produção das vogais nasais. A baixa resolução temporal, a aquisição dos segmentos nasais durante produções sustentadas e não em fala encadeada e a impossibilidade de obter um sinal acústico de qualidade são apontadas como as maiores limitações do estudo.

O outro estudo de RM em tempo real encontrado para o PB foi conduzido por Medeiros e Demolin (2006) com o objetivo de observar os movimentos de abaixamento do véu palatino durante a produção de vogais nasais e correspondentes vogais orais do PB. O estudo incluiu um falante do género masculino, de S. Paulo com formação académica superior. No *corpus*, foram considerados cinco pares mínimos contendo as vogais [i, e, a, o, u], com distinção de significado pelo traço oral *versus* nasal (e.g., **mito/minto**). O contexto à direita da vogal nasal não foi tido em consideração. As aquisições de imagem decorreram no Hospital *Erasmus* em Bruxelas seguindo o protocolo desenvolvido por Demolin *et al.* (2000). A sequência utilizada foi o TSE Zoom, ponderação T1 no plano sagital, espessura de corte de 6mm, permitindo obter 5 imagens/s. Não foi efetuada gravação do sinal acústico.

Os dados apresentados são de natureza qualitativa, não tendo sido levado a cabo qualquer tipo de segmentação ou processamento sobre as imagens. Os resultados obtidos com o estudo são, de acordo com os autores, *“assumidamente preliminares e pouco conclusivos quanto às questões da nasalidade no PB”*. De facto, os resultados resumem-se à constatação que o véu palatino está elevado na produção das vogais orais e descido na produção das nasais (podendo variar de vogal para vogal). É referido pelos autores que a posição da língua, entre vogais orais e nasais, é

diferente e que essas diferenças se devem a mecanismos de compensação para criar configurações de ressonância do trato adequadas à qualidade vocálica desejada. Os autores referem ainda a questão da duração dos segmentos vocálicos nasais, mais longos que do que das correspondentes orais. Constatam ainda que, apesar de a resolução temporal ser muito baixa, não permitindo detetar o intervalo de tempo relativo ao início da vogal, existe, pelo menos, em todas as vogais nasais uma imagem em que o véu palatino está descido, o que os leva a defender a existência de uma fase nasal da vogal, de duração longa, que será responsável pela distinção nasal/oral.

Existem ainda dois estudos que, apesar de não incluírem vogais nasais, devem ser aqui mencionados, uma vez que se focalizam no estudo dos aspetos temporais e de coordenação entre gestos para as consoantes nasais do Inglês, que utilizam RM em tempo real. Referimo-nos aos estudos de Byrd *et al.* (2009) e de Kim *et al.* (2010).

Um dos objetivos do estudo apresentado por Byrd *et al.* (2009) foi precisamente verificar a adequabilidade da utilização da RM em tempo real para o estudo dos padrões de coordenação temporal entre gestos, na produção de segmentos nasais, como o [n]. Estes sons envolvem um gesto velar (abaixamento) e uma constrição no trato oral.

No estudo, participaram quatro falantes (três homens e uma mulher) todos falantes de Inglês Americano, mas provenientes de diferentes estados. Como o padrão de coordenação pode variar com a estrutura silábica e com o acento, estes aspetos foram contemplados no *corpus* adquirido. Um dos conjuntos de estímulos foi constituído por pares de palavras monossilábicas com a consoante [n] em posição inicial, final e intervocálica. O outro conjunto de estímulos incluiu palavras com a nasal em posição interna, mas com diferente acentuação.

As imagens foram obtidas recorrendo ao protocolo de tempo real proposto por Narayanan *et al.* (2004), utilizando uma sequência EG com preenchimento espiral do espaço K com resolução temporal de 11 *frames/s* (informação a cada 89 *ms*). A informação original foi reconstruída com utilização de *sliding windows*, permitindo obter uma *frame rate* de 21 *frames/s*. Foi efetuada gravação do sinal acústico, em simultâneo, recorrendo a tecnologia de fibra óptica.

Para cada uma das imagens adquiridas foram obtidos os valores de abertura mínima do trato oral (para o gesto da ponta da língua) e a distância entre a faringe e o véu palatino para o gesto velar. O máximo valor de abertura velar e o mínimo de abertura ao nível da ponta da língua (*extremum*) foram determinados a partir da derivada das curvas de abertura, nas passagens por zero. Os autores constataram: i) a existência de um padrão temporal bimodal, no qual se verifica sincronia dos gestos em posição inicial e padrão sequencial em coda; ii) que as nasais em posição intervocálica apresentam padrões de coordenação que são sensíveis ao acento; iii) que a resolução temporal conseguida com RM adequa-se ao estudo dos aspetos temporais e de coordenação, neste caso com aplicação às consoantes nasais (Byrd *et al.*, 2009).

O estudo de Kim *et al.* (2010) tem características mais técnicas e teve como objetivo testar um método de reconstrução retrospectiva (*spiral golden ratio*), que permitiu obter imagens com menos artefactos do que o método habitualmente utilizado com as aquisições espirais. Para além disso, uma vez que os diferentes articuladores (e.g., ponta da língua e véu palatino) se movem com diferente velocidade, os autores efetuam a reconstrução da informação com diferente resolução

temporal.

A sequência utilizada foi, também, a proposta por Narayanan *et al.* (2004). Participou um sujeito do gênero masculino e foram utilizados 6 estímulos com a consoante nasal [n] em diferentes posições silábicas. Obtiveram os perfis de intensidade (ao longo do tempo) relativos à abertura do véu palatino e ponta da língua, utilizando quatro janelas temporais diferentes, conduzindo a resolução temporal variável: 48 ms, 78 ms, 126 ms e 204 ms. Verificaram que os movimentos da ponta da língua são melhor visualizados (menos artefactos) com uma resolução temporal de 48 ms e que o véu palatino é observado claramente, sem artefactos e com S/R adequada, com uma resolução temporal de 126 ms. Os autores argumentam que a seleção flexível da resolução temporal em conjunto com o método de reconstrução proposto permitirão caracterizar melhor os aspetos de coordenação temporal do véu palatino e dos articuladores orais, na produção dos sons nasais. Na nossa opinião este estudo é muito importante no sentido em que reforça o conceito de que não há uma resolução temporal única e absoluta que se deve atingir para conseguir estudar a fala. Esta deve ser adequada ao objetivo concreto de cada estudo, em função da informação que se pretende obter, dos sons em análise e das suas características mais importantes. Parece-nos óbvio que ao melhorar a resolução temporal vai-se melhorando a capacidade de caracterizar ou discriminar eventos mais rápidos, contudo, muitas vezes perde-se informação espacial que também pode ser relevante para a obtenção da informação desejada. Esta filosofia de pensamento terá que estar subjacente quando se estabelece o protocolo de aquisição.

Será oportuno ainda fazer um pequena sùmula de outros aspetos relevantes relacionados com a utilização de RM em tempo real no contexto da produção de fala. No que diz respeito ao número de falantes, verifica-se que este é, no geral, bastante pequeno variando entre um a quatro falantes. Podemos constatar que houve evoluções importantes nestes estudos, nomeadamente: i) melhoria da resolução temporal verdadeira, obtenção de *frame rate* elevado e possibilidade de reconstruir a informação com diferente resolução temporal; ii) crescente utilização de MO para gravação do sinal acústico iii) apesar de ainda se recorrer muito a técnicas de segmentação manual ou, como referido, não ser efetuado qualquer tipo de processamento sobre as imagens tem havido um grande investimento na área do processamento de imagem e dos dados, possibilitando a obtenção de informação quantitativa de forma bastante rápida e facilitada.

Em consequência, a RM torna-se cada vez mais aliciante nesta área permitindo a aquisição de *corpora* mais extensos, adquirir palavras e frases em contexto, ao invés de sons isolados, acompanhar os movimentos dos articuladores ao longo do tempo, obter informação relativa aos padrões de coordenação entre gestos, sendo possível alargar o leque de sons e de aspetos que podem ser estudados utilizando esta modalidade imagiológica. Para além disso, os resultados já obtidos com RM em tempo real apontam claramente para o grande potencial da RM na obtenção de informação comparável e complementar aos dados EMA, cine-radiografia ou *velotrace* realizados no passado (Delvaux *et al.*, 2002a; Byrd *et al.*, 2009). Tendo em consideração as características dinâmicas da nasalidade vocálica no PE, a obtenção de informação temporal é extremamente relevante.

6.3 Justificação e objetivos do estudo

Ao contrário dos sons laterais, em que a inexistência de dados articulatórios para o PE constituiu uma das principais motivações para o nosso estudo, para os sons nasais, particularmente para as vogais nasais do PE, existem já alguns dados articulatórios, obtidos através de EMA (Teixeira e Vaz, 2001; Oliveira e Teixeira, 2007b; Oliveira, 2009) e de RM estática (Martins, 2007; Oliveira, 2009). Os dados obtidos nessas investigações foram um importante avanço relativamente ao conhecimento acerca da produção das vogais nasais, no entanto apresentam algumas limitações que passamos a enumerar:

- Apesar da resolução temporal conseguida com EMA ser bastante elevada, a informação obtida: i) é limitada a um pequeno conjunto de pontos (relativos aos sensores) que reúnem informação dos vários articuladores; ii) provém de um único sensor no caso do véu palatino; iii) é condicionada pelo facto de nem sempre ser possível colocar os sensores exatamente no mesmo local, para os diferentes falantes;
- Os dados EMA foram recolhidos (Teixeira e Vaz, 2001; Oliveira, 2009) em alturas e com equipamentos EMA diferentes. A primeira base de dados foi adquirida para um falante e a segunda para dois falantes;
- Os resultados dos dois estudos EMA nem sempre foram concordantes, nomeadamente no que diz respeito à posição do véu palatino na produção das diferentes vogais nasais do PE, dificultando generalizações;
- Os dados de RM recolhidos até ao momento permitem obter informação de todo o trato vocal, e não só de pontos, mas basearam-se em aquisições estáticas (2D e 3D) durante a produção sustentada dos sons e, desta forma, só fornecem informação relativa a um determinado momento da produção da vogal nasal, não refletindo o seu carácter dinâmico.

Estas condicionantes justificam que se retome o estudo das vogais nasais, desta vez com RM em tempo real, esperando obter informação dinâmica não restrita a um conjunto de pontos e extrair novos dados de produção só possíveis de obter, de forma direta, através de RM (e.g., áreas na zona de acoplamento), com o objetivo de poder contribuir para um maior conhecimento da nasalidade vocálica para o PE, que viabilizará também melhorias na qualidade dos sons sintetizados.

Tendo em mente o exposto nos parágrafos anteriores, considerando que não temos ao nosso dispor protocolos ajustados e otimizados, para aquisição de imagem em tempo real, nem experiência com o processamento deste tipo de imagens, os objetivos delineados para esta parte do nosso trabalho contemplam aspetos de carácter técnico, que passamos a enumerar:

- Otimização de protocolos RM com obtenção de imagens em tempo real, com resolução temporal adequada, que permitam captar as características dinâmicas de sons nasais durante a produção de fala;
- Implementação de procedimentos que viabilizem a aquisição sincronizada do sinal acústico durante a aquisição de imagens de RM;

- Explorar as potencialidades da RM, nomeadamente a sua capacidade multiplanar e versatilidade, no sentido de obter novos dados de produção (e.g., variação de áreas no tempo) que possam contribuir para a modelação das vogais nasais;
- Utilizar as imagens obtidas para testar várias abordagens de processamento, utilizando rotinas de segmentação com o mínimo de interação humana, que permitam análise sistemática e quantitativa dos dados obtidos;
- Explorar metodologias de análise dos dados, que permitam obter informação semelhante à obtida previamente com EMA.

A consecução destes objetivos viabilizará o retomar de algumas questões, ainda não totalmente esclarecidas (veja-se a sua formulação original em Teixeira e Vaz, 2001), relativas à produção das vogais nasais para o PE, nomeadamente:

1. De que forma variam, ao longo do tempo, os movimentos do véu palatino e qual a duração da abertura e do fecho?
2. De que forma varia a posição da língua, ao longo da produção de uma vogal nasal?
3. Existirão ou não ajustes orais durante a produção da vogal nasal?
4. Como variam ao longo do tempo as áreas nasais e orais durante a produção de uma vogal nasal? Qual a magnitude dessas áreas e sua covariação?
5. Como se coordenam temporalmente os diferentes articuladores entre si?
6. Existirão, de facto, três fases distintas (oral, nasal e segmento consonântico) durante a produção de uma vogal nasal?
7. Quais as diferenças entre uma vogal produzida de forma isolada e em contexto?

6.4 Método

Nesta secção, começaremos por descrever o *corpus* de sons da fala elaborado para esta experiência, seguida pela caracterização dos falantes que estiveram envolvidos na recolha de dados. Dar-se-á conta dos procedimentos técnicos (e.g., equipamentos, protocolo) que envolveram as recolhas de dados (RM e acústicos). Por fim, descrevem-se, sucintamente, as técnicas de processamento de imagem e de tratamento de dados que foram utilizadas neste estudo e que viabilizaram a apresentação dos resultados.

6.4.1 *Corpus*

O *corpus* elaborado, para as aquisições em tempo real, enquadra-se nos objetivos delineados para o projeto HERON II, visando fundamentalmente o estudo das vogais nasais. Os *corpora* já recolhidos com EMA para o PE (Teixeira e Vaz, 2001; Oliveira, 2009) serviram de guia para a

sua elaboração⁴, no entanto foram necessárias adaptações incluindo a seleção de apenas alguns contextos, ponderando as questões específicas que se pretendem vir a abordar, as condicionantes técnicas e, não menos importante, um tempo de aquisição total que fosse razoavelmente tolerado pelos participantes.

Assim, o *corpus* final contemplou as seguintes situações:

1. As cinco vogais nasais e correspondentes vogais orais do PE, adquiridas de forma isolada;
2. Vogais nasais nas três posições lexicais (inicial, medial e final) como, por exemplo, **ampa**, **pampa** e **pan**. Apenas foi considerado o contexto com oclusiva bilabial [p] e fricativa labial [f];
3. Vogal nasal e vogal oral, em contexto simétrico, com oclusiva bilabial como **pampa** [p6 ~p6] e **papa** [pap6].

As vogais nasais e orais isoladas foram adquiridas pedindo ao falante que produzisse, sequencialmente, as vogais durante o período de aquisição. As vogais orais foram sub-divididas em abertas e fechadas para facilitar a tarefa ao falante (e.g., sequência de vogais orais abertas - [a, E, i, O, u]). Cada sequência foi repetida três vezes. Para cada uma das vogais nasais, nas posições silábicas consideradas, foram efetuadas três aquisições.

As vogais orais, em contexto de oclusiva bilabial [p], foram produzidas em séries diferentes, em que o falante repetiu quatro a cinco vezes a palavra com a vogal em causa (e.g., **pipa**, **pipa**, **pipa**) durante o tempo de aquisição. Os falantes foram instruídos para produzirem cada uma das palavras ou sequências do *corpus* com entoação normal e a um ritmo confortável. Com exceção de alguns casos selecionados, as palavras não foram inseridas em frase de suporte para não prolongar o tempo total de aquisição de imagem. Quando utilizada, a frase de suporte foi: “Diz **ampa**, **pampa**, **pan** após”. Foram também adquiridas consoantes nasais e, de forma complementar e exploratória, algumas palavras contendo consoantes laterais e róticos, com vista a planear e otimizar futuras aquisições de dados. No anexo G apresenta-se a descrição detalhada do *corpus*.

6.4.2 Falantes

Neste estudo, participaram três falantes nativos do Português Europeu (PE), todos do sexo feminino. O número de falantes é considerável quando comparado com os estudos similares, já referidos. Apresentamos a sua caracterização na altura das recolhas de dados:

- Falante CO, 33 anos, natural de Viseu, residente na cidade de Aveiro há 7 anos. Habilitações académicas a nível do Doutoramento. Professora do Ensino Superior. Tem conhecimentos de fonética e linguística. Mede 160 *cm* e pesa cerca de 50 *Kg*. Esta falante participou nas aquisições 3D relativas aos sons laterais.

⁴O *corpus* adquirido no estudo RM, para o francês, efetuado por Delvaux *et al.* (2002a) serviu também de referência para a delineação do *corpus*.

- Falante CM, 21 anos, natural de Seia (distrito da Guarda), residente na cidade de Aveiro há 3 anos. Estudante do curso de Terapia da Fala, tem conhecimentos de fonética. Mede 161 *cm* e pesa 52 *Kg*.
- Falante SV, 21 anos, nasceu no Canadá, filha de pais portugueses, veio para Portugal (zona de Fátima) com 3 anos. A língua materna é o português, não adquiriu uma segunda língua. Reside em Aveiro há 3 anos. Estudante do curso de Terapia da Fala, tem conhecimentos de fonética. Mede 174 *cm* e pesa 60 *kg*.

Os critérios de inclusão e exclusão seguidos foram os que já apresentámos no capítulo precedente, para as recolhas do *corpus* RM 3D. De forma similar, foram recolhidos alguns dados sobre as falantes (anexo B), foi obtido consentimento informado (anexo C) e preenchido o questionário de rastreio de RM (anexo D) utilizado na instituição onde as aquisições foram efetuadas, dando cumprimento aos requisitos éticos (CIOMS, 2002) e de segurança (ICNIRP, 2004; Sherlock, 2001; Sherlock e Spinazzi, 2008) inerentes à presente investigação. Todas as falantes participaram de forma voluntária e gratuita.

6.4.3 Recolha de dados

6.4.3.1 Local e Período

A aquisição das imagens de RM foi efetuada através da Rede de Imagiologia Funcional, no IBILI, em Coimbra. No período de maio a setembro de 2011, foram efetuadas algumas sessões experimentais. A recolha efetiva dos dados de RM em tempo real, para as três falantes, foi efetuada em duas sessões separadas, com cerca de 60 a 75 minutos, por falante. As sessões decorreram no mês de outubro de 2011 (falante CO) e fevereiro de 2012 (falantes SV e CM) com o suporte técnico considerado necessário e fornecido pela instituição.

6.4.3.2 Equipamento e material

As imagens de RM foram adquiridas num equipamento com intensidade de campo de 3T (Magnetom TIM TRIO, Siemens, Erlangen, Alemanha) equipado com gradientes de elevada *performance* (Máxima amplitude - 45 *mT/m*; *Rise time* - 200 μ s; *Slew Rate* - 200 *T/m/s*; FOV-50 *cm*). Foram utilizadas, em simultâneo, antenas *phased-array* de crânio (12 canais) e de pescoço (4 canais) com recurso a imagem paralela (GRAPPA). A antena de corpo, incorporada no sistema, foi utilizada como antena transmissora.

Para a gravação do sinal acústico, em simultâneo, com as imagens de RM foi utilizado um microfone ótico, da *Optoacoustics*, modelo *FOMRI III Dual channel* (Or Yehuda, ISRAEL). Este sistema e modelo constitui a mais recente tecnologia nesta área, podendo ser utilizado para gravação de som, durante a aquisição de imagens de RM, com sistemas de elevado campo magnético (até 7T), apresentando elevados níveis de desempenho (Optoacoustics, 2010).

6.4.3.3 Sessões Experimentais

De forma semelhante ao efetuado para as aquisições 3D, houve necessidade de um período preparatório que antecedeu a recolha de dados com o intuito de:

- escolher e otimizar as sequências de pulso;
- testar o microfone ótico e, posteriormente efetuar testes para otimizar a sincronização das imagens com as gravações de som efetuadas;
- otimizar procedimentos para tornar mais rápidas e produtivas as sessões de recolha de dados e minimizar possíveis erros (e.g., codificação das imagens) que possam vir a comprometer, à *posteriori*, a exploração dos dados.

As experiências realizadas durante este período são de extrema importância uma vez que daí resultou o protocolo final com que foi adquirido o *corpus* e permitiram também testar uma série de procedimentos que têm que estar perfeitamente afinados para as sessões de aquisição. O trabalho de otimização constitui um dos objetivos do presente trabalho. No entanto, como será necessário conferir algum detalhe, a sua descrição poderia quebrar um pouco a harmonia do presente capítulo. Desta forma, optou-se por criar um anexo a este capítulo que contempla estes aspetos (veja-se anexo F).

6.4.3.4 Protocolo de aquisição final

Apesar de todo o trabalho preparatório, no dia das aquisições é necessário, para cada um dos falantes, fazer um pequeno ensaio com o MO. Proporciona-se também ao falante uma entrada dentro da *gantry* para se familiarizar com o ambiente e para testar alguns procedimentos (e.g., forma de comunicação durante a aquisição de imagem) sendo esclarecida qualquer dúvida que o falante coloque.

Posteriormente, o falante é posicionado em decúbito dorsal com os joelhos apoiados, por forma a ficar o mais confortável possível. São colocadas as antenas, imobilizada a cabeça com apoios adequados e colocado o MO. Este tem que ser montado na antena de crânio e ajustado de forma muito precisa por forma a ficar a cerca de 1 *cm* dos lábios do falante mas sem lhes tocar. Junto com o MO são fornecidas proteções para colocar no microfone (*pop-screens*) que têm utilização única. A figura 6.1 ilustra o posicionamento do falante e a colocação do MO na antena. Nesta altura, antes de ser colocado dentro da *gantry* solicita-se ao falante para efetuar uma produção, que será gravada com o MO, sem qualquer ruído (seguindo as instruções do fabricante). Depois disso efetua-se a centragem sendo posicionado dentro da *gantry* para iniciar a sessão.

Primeiro, procedeu-se à aquisição das imagens de localização (*localizer*) e de seguida efetuou-se a marcação de um corte sagital na linha média com 8 *mm* de espessura, FOV de 210 *mm* por forma a englobar todo o trato vocal e prevenir artefactos de *aliasing* que, neste caso, terão probabilidade de ocorrer na direção antero-posterior (AP) uma vez que é esta a direção em que é efetuada a codificação em fase. Depois de obtida uma série de teste, inicia-se a aquisição

de todos os itens contemplados no *corpus*, no plano médio-sagital, pela ordem pré estabelecida na folha registo. A tabela 6.1 apresenta os parâmetros da sequência otimizada e utilizada para as aquisições de imagem em tempo real (coluna referente a 14 *frames/s*)⁵.

Antes de se fazer a aquisição de uma série o falante é informado, através do intercomunicador, qual é a palavra ou sequência de palavras que deve produzir. De seguida, inicia-se a gravação com o MO (frequência de amostragem de 16 *kHz*) e é lançada a sequência de pulsos. A partir do momento que começa a ouvir o ruído produzido pela máquina o falante aguarda entre 1 a 2 s e só depois inicia a produção⁶ (veja-se anexo F, para mais detalhes). Sempre que o falante teve dúvida relativamente à produção foi efetuada uma repetição. Na consola, após cada série, o sinal áudio e as imagens foram sempre conferidos, de forma rápida, antes de se avançar para nova aquisição.



Figura 6.1: Posicionamento do falante e colocação do MO na bobine de crânio. Imagem obtida na primeira sessão de aquisição com falante a CO.

Para além do corte sagital realizado na linha média, para alguns itens do *corpus* (e.g., vogais nasais nas diferentes posições lexicais em contexto de oclusiva bilabial), foi efetuado um corte coronal oblíquo na zona do véu palatino. Estes foram adquiridos após a realização de todas as séries no plano sagital. Para efetuar a marcação do corte recorreu-se ao corte sagital correspondente à mesma produção a adquirir no plano coronal. A orientação do corte foi estabelecida tendo por referência o mento e o “joelho” do véu palatino correspondendo a uma orientação de aproximadamente 40° com o plano transversal. A escolha desta orientação não é aleatória exis-

⁵Esta foi a resolução temporal selecionada para as aquisições. Existe já protocolo otimizado com resolução temporal superior - 23 *frames/s* apresentando-se na tabela os parâmetros utilizados.

⁶Os falantes tiveram acesso prévio ao *corpus* para estarem familiarizados com os sons e com a forma como iriam ser produzidos. Foi também explicada o processo como se iria estabelecer a comunicação durante a aquisição. Neste caso como todos tinham conhecimentos de fonética a comunicação tornou-se mais simples.

Tabela 6.1: *Parâmetros de aquisição utilizados, com sequência TurboFLASH, com diferente resolução temporal.*

Parâmetro	14 <i>frames/s</i>	23 <i>frames/s</i>
Sequência	<i>TurboFLASH</i>	<i>TurboFLASH</i>
Modo	single shot 2d	single shot 2d
TR	72 <i>ms</i>	44 <i>ms</i>
TE	1.02 <i>ms</i>	0.82 <i>ms</i>
FA	5	5
ETL	1	1
FOV _{leitura}	210	210
FOV _{fase}	100%	100%
Intervalo	NA	NA
Orientação	Sagital/coronal oblíquo	Sagital
Esp. corte	8 <i>mm</i>	8/10 <i>mm</i>
Resolução _{base}	128	64
Resolução _{fase}	50%	100%
RecFOV	6/8	6/8
Passos de codificação fase	48	48
PAT	GRAPPA	GRAPPA
Fator de aceleração _{PE}	2	4
Linhas de referência _{PE}	24	17
Número de excitações (NEX)	1	1
Tempo de aquisição	5.4 <i>s</i> (75 imagens)	6.7 <i>s</i> (150 imagens)
Largura banda (BW)	1395 <i>Hz/pixel</i>	1395 <i>Hz/pixel</i>
Espaçamento entre ecos	2.0 <i>ms</i>	1.8 <i>ms</i>
Espaçamento pixel	0.82 × 0.82	0.82 × 0.82
Modo de gradientes	Rápido	Rápido

tindo alguns factos na literatura consultada que consubstanciam esta opção⁷. Os cortes coronais oblíquos oferecem a possibilidade de medir a área nasal ao longo do eixo em que se movimenta o véu palatino e observar a sua variação no tempo. Para além disso é possível obter a distância médio-sagital, a posição do véu palatino e obter ainda a área oral. Pode-se, desta forma, relacionar a altura e abertura do véu palatino, ao longo do tempo, de forma simplificada.

Com o intuito de poder obter informação temporal, em simultâneo, a nível do véu palatino e dos lábios, foi possível a aquisição de dois cortes coronais com sacrifício da resolução temporal. A figura 6.2 apresenta a orientação escolhida para cada um dos cortes efetuados. Neste caso, o número de *frames* baixou para sete o que corresponde a informação a cada 144 *ms*. Os cortes são adquiridos “virtualmente” em simultâneo. Na realidade, estão desfasados por um período correspondente a 72 *ms*. Para se conseguir obter informação dos dois articuladores no mesmo ponto temporal terá que, posteriormente, ser efetuada interpolação de uma ou das duas séries

⁷A realização de cortes coronais na zona do véu palatino permitirá revisitar, utilizando RM, alguns temas que nas décadas de 70 e 80 ocuparam muitos investigadores recorrendo a técnicas como o *Velotrace* (e.g., Bell-Berti, 1993) ou fibroscopia (e.g., Benguerel *et al.*, 1977) para obter dados relativamente aos movimentos do véu palatino ao longo do tempo, na produção de sons nasais e não nasais. O *velotrace* e a fibroscopia permitem, através da introdução de dispositivos na narina observar o véu palatino num plano frontal e obter de forma direta a posição do véu palatino (altura) e, em simultâneo, obter a área transversal na zona do véu palatino. Estas técnicas, mais invasivas, mas que não utilizavam radiação ionizante, permitem complementar a informação obtida através de cine-radiografia onde as áreas eram estimadas, a partir da distância médio sagital, mas não diretamente medidas (Moll e Daniloff, 1971; Shaw *et al.*, 1980). Nestes estudos o movimento do véu palatino era descrito tendo por base uma orientação oblíqua, tendo como referência o “joelho” do véu palatino. Mais recentemente, Serrurier e Badin (2008) verificaram que é o movimento do véu palatino ao longo desta direção que explica cerca de 83% da sua variação.

de cortes. A figura 6.3 mostra um exemplo de como se poderá fazer o alinhamento das *frames*. Note-se que as *frames* a cinzento são interpoladas.

Na tabela 6.2 apresenta-se um resumo da resolução temporal e resolução espacial obtidas para as diversas séries.

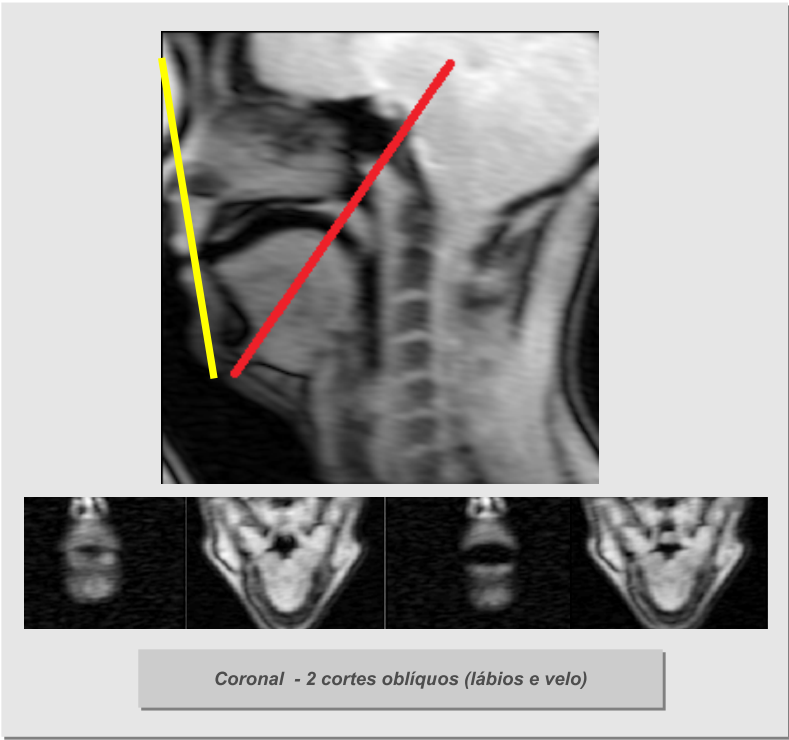


Figura 6.2: Em cima, orientação escolhida para a marcação dos cortes coronais nos lábios (linha amarela) e no véu palatino (linha vermelha). Em baixo, exemplo das imagens obtidas: resolução temporal de 144 *ms*.

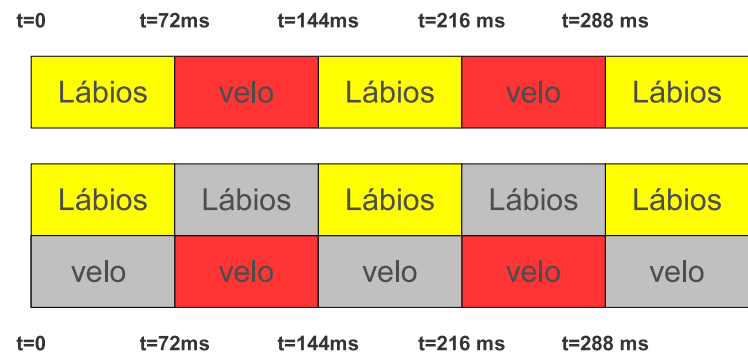


Figura 6.3: Exemplo ilustrativo da forma como se poderá obter informação sincronizada para os dois articuladores, através de um processo de interpolação linear.

Depois de adquirido todo o material relativo aos sons nasais (a prioridade), e depois de confirmar com o falante a sua condição e disponibilidade para avançar adquiriram-se algumas séries com sons laterais e róticos e efetuaram-se mais algumas experiências (material complementar)

Tabela 6.2: Tabela resumo com os valores relativos à resolução temporal, resolução espacial e frames/s conseguidas para as várias séries.

Orientação	N ^o cortes	Res. Temp (ms)	Esp. corte (mm)	Matriz	Pixel (mm ²)	Frames/s
Sagital	1	72	8	64 × 128	3.3 × 1.6	14
Sagital	1	44	8	64 × 64	3.3 × 3.3	23
Coronal	1	72	8	64 × 128	3.3 × 1.6	14
Coronal	2	144	8	64 × 128	3.3 × 1.6	7

como se pode observar no anexo G. Saliente-se que todos os falantes participantes nesta experiência adquiriram a totalidade do *corpus* previsto e nenhum foi reposicionado durante a sessão de aquisição. Apesar de cada uma das aquisições representar apenas segundos são efetuadas cerca de 70 séries, cada uma das quais demora, se tudo estiver otimizado, cerca de 1 minuto a preparar.

6.4.4 Processamento do sinal de voz

A primeira tarefa que teve que ser concretizada para viabilizar a extração de informação foi o tratamento dos dados acústicos. Para fazer o alinhamento da imagem com o áudio utilizou-se como referência o primeiro da série de pulsos TTL⁸ emitidos através da sequência de pulsos (veja-se anexo F). Este pulso é coincidente com a primeira imagem da série. Os ficheiros foram cortados e renomeados seguindo codificação idêntica à utilizada para identificar as diferentes séries de imagens.

Posteriormente, foram anotados, manualmente, com recurso ao programa Praat (Borsma e Weenink, 2009). Para efetuar a anotação fonética recorreu-se ao alfabeto X-SAMPA⁹. Desta forma torna-se possível identificar o conjunto de imagens que correspondem à produção de cada um dos diferentes fonemas produzidos. Um exemplo de anotação do ficheiro áudio, para uma das produções da falante CO, pode ser observado na figura 6.4.

A parte inicial do espectrograma corresponde ao período de tempo em que já se está a adquirir imagens e o falante ainda não iniciou a produção. A aquisição tem que ser efetuada desta forma pois o MO necessita de um pequeno período de tempo (menos de 1s) para conseguir efetuar o cancelamento do ruído produzido pelas bobines (zona assinalada a vermelho, à esquerda, na figura 6.4). Como se pode constatar no espectrograma, apesar do sistema de cancelamento de ruído do MO, observa-se algum ruído de fundo que impede uma análise acústica muito detalhada.

6.4.5 Processamento de Imagem

O processamento de imagem foi efetuado no âmbito de uma das tarefas do Projeto HERON II, em que a autora é também investigadora, sendo muito do seu desenvolvimento da responsabilidade do Doutor Samuel Silva. Apesar do envolvimento da autora em todas as etapas do processo, nomeadamente nas decisões quanto ao tipo de informação a extrair e possíveis abordagens a efetuar,

⁸Transistor-Transistor Logic.

⁹Agradecemos à Doutora Catarina Oliveira o trabalho de anotação dos ficheiros áudio. A realização desta tarefa exige conhecimentos de fonética e acústica que vão para além do domínio de conhecimentos da autora do presente trabalho.

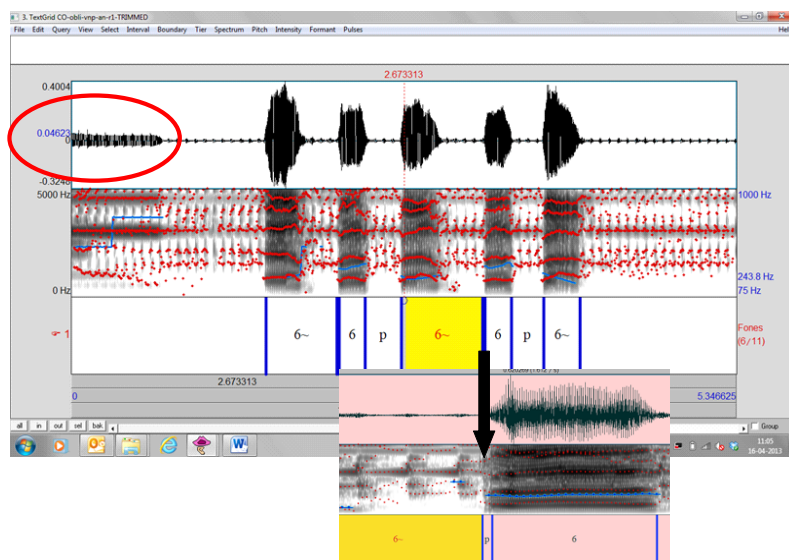


Figura 6.4: Sinal acústico, espectrograma e respetiva anotação fonética para a sequência **ampa, pampa, pan** para uma das produções da falante CO

as tarefas de processamento de imagem não foram efetuadas por si. Desta forma, apresentar-se-á apenas uma pequena síntese das metodologias utilizadas e que viabilizaram a obtenção de alguns resultados. A informação extraída das imagens de RM é variada e resulta quer das imagens sagitais quer das imagens coronais adquiridas. A partir das imagens sagitais foram extraídos contornos 2D e as curvas de variação do véu palatino e dos lábios ao longo do tempo. Dos cortes coronais obteve-se a variação das áreas das cavidades oral e nasal ao longo do tempo e, a partir destas, calculou-se a proporção nasal/oral (P_{NO}); os valores da duração dos movimentos de abertura/fecho do véu palatino foram também obtidos a partir dos cortes coronais. No seu conjunto os dados permitem caracterizar os diversos articuladores envolvidos na produção das vogais nasais.

6.4.5.1 Extração de contornos

Dos cortes obtidos no plano médio-sagital foram extraídos os contornos do trato vocal usando *Region Growing* (RG), numa implementação efetuada com o *MeVisLab*. A segmentação tem início com a definição da *Region of Interest* (ROI) ou região de interesse que engloba todo o trato vocal. A ROI é definida numa das imagens da sequência temporal e é replicada para toda a série. O objetivo é evitar que o algoritmo de segmentação evolua para zonas que não são de interesse (e.g., para lá dos lábios, cavidade nasal ou palato duro). O RG é depois iniciado a partir de um ponto semente colocado dentro da ROI definida. Devido à existência de coerência espacial, se não existirem descontinuidades, uma só semente será necessária para segmentar toda a série de imagens;

Antes de iniciar a segmentação propriamente dita foi efetuada uma interpolação linear permitindo obter pontos temporais mais próximos. Assim, para cada série o número de imagens a segmentar passou de 75 para 150 imagens (com informação a cada 36 ms).

Para melhorar a segmentação a nível dos lábios foi criada uma segunda ROI e utilizados parâmetros de intensidade otimizados para esta região. Como não houve movimento dos falantes nem reposicionamento dos mesmos durante toda a sessão de aquisição, as ROI utilizadas para os lábios e trato vocal bem como os valores de intensidade e localização dos pontos semente aplicam-se a todas as produções. Desta forma, as segmentações são efetuadas de uma forma rápida com pouca interação do operador. Contudo, se houver necessidade a segmentação pode ser editada para acrescentar ou remover zonas do trato vocal. Na figura 6.5 apresentam-se duas situações em que foi necessário editar e efetuar correções na segmentação proposta pelo algoritmo.

A segmentação de toda a série demora entre 1 a 5 minutos, dependendo de pequenas correções que possam ter que ser efetuadas e contempla, também, o tempo despendido para a revisão global das segmentações em todas as *frames* (Silva *et al.*, 2013a).

Extraídos os contornos 2D para todas as *frames*, e identificadas a *frame* ou *frames* de interesse, com base no alinhamento com o ficheiro áudio, estão criadas as condições para efetuar sobreposições de diferentes contornos, em função do que se pretenda estudar. Por exemplo, comparar o mesmo som em diferentes contextos/posições silábicas ou obter a informação sobre a variação dos articuladores no tempo (informação dinâmica).

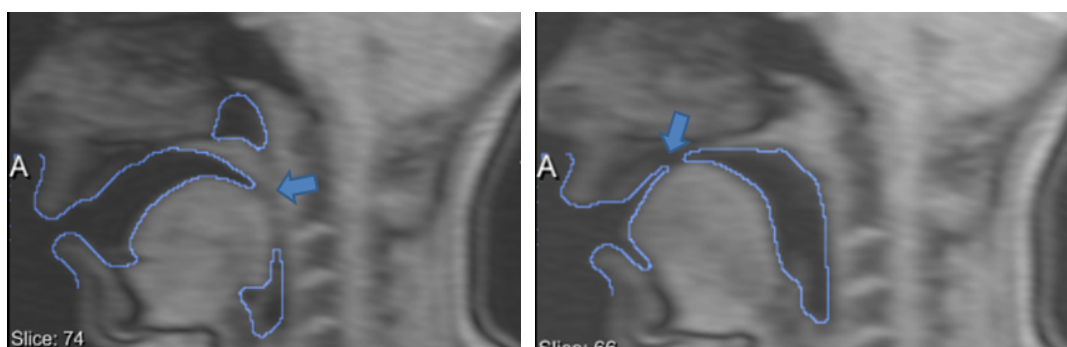


Figura 6.5: Exemplo de segmentações que necessitam de edição e correção por parte do operador

As comparações através de sobreposição de contornos são facilitadas se não existir movimento dos falantes ao longo da sessão de aquisição de imagem e essa será a situação ideal. Se isso se verificar, antes de se efetuar as comparações poderá ter que se efetuar re-alinhamento utilizando para isso um ponto de referência anatómico em que não seja espetável a existência de alterações (e.g., uma das vértebra cervicais). Quando são efetuadas várias repetições do mesmo som, poderá, como é sabido, existir alguma variabilidade. A figura 6.6, apresenta a sobreposição de contornos para três repetições do mesmo som para cada uma das falantes, permitindo constatar que não existiu movimento significativo das falantes e, ao mesmo tempo, houve uma grande consistência nas diferentes produções efetuadas por cada uma das delas. Note-se que cada uma das produções foi efetuada em diferentes séries.

Na figura 6.7 apresenta-se, a título ilustrativo, um exemplo relativo à falante CO na produção de **empa** [e~p6]. Os contornos à esquerda referem-se à vogal nasal [e~] e à direita à vogal oral [6]. Com a sobreposição de contornos obtidos em diferentes pontos temporais ao longo da produção do som consegue-se observar quais os articuladores que sofrem alterações e a direção

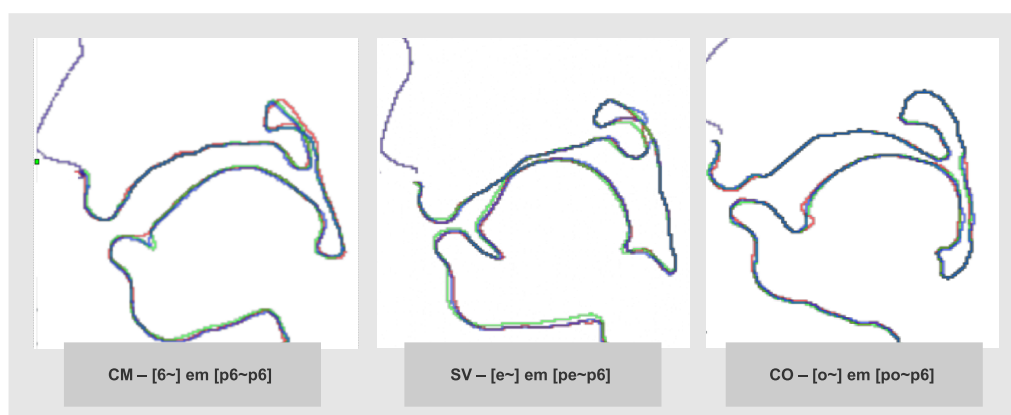


Figura 6.6: Sobreposição de contornos relativos à mesma vogal para cada uma das três repetições efetuadas com vogal nasal em posição medial no contexto de oclusiva bilabial. Falante CM - [ẽ~] (à esquerda), falante SV - [ẽ~] (ao centro) e falante CO - [o~] (à direita).

desses movimentos (e.g., recuo/avanço do dorso da língua, posição mais ou menos descida do véu palatino, etc). Note-se que para a vogal oral não há grandes alterações ao longo do tempo. Os resultados obtidos mediante a comparação de contornos apresentam-se na secção 6.5.2.

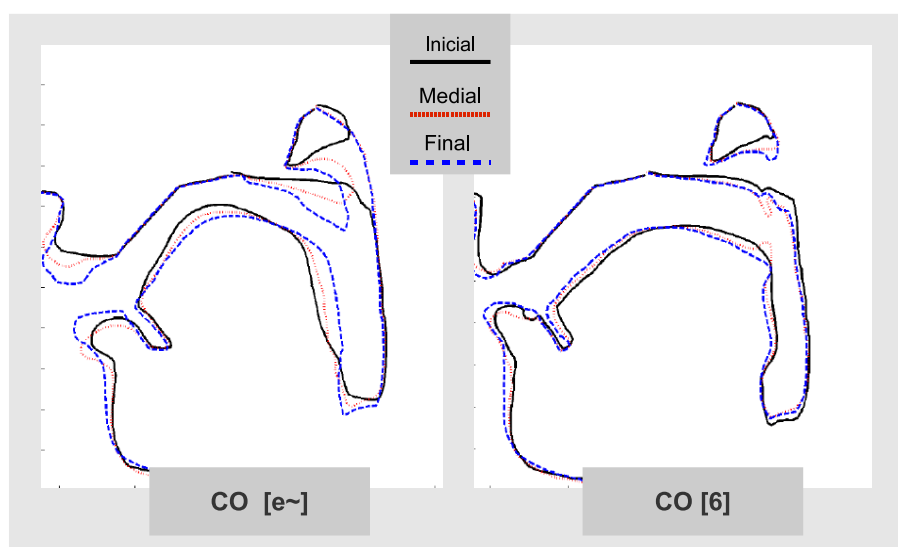


Figura 6.7: Exemplo de sobreposição de contornos 2D, para a falante CO. À esquerda representados três contornos ao longo da produção de uma vogal nasal [ẽ~] e à direita para a vogal oral [6].

6.4.5.2 Extração de informação relativa ao movimento e coordenação dos articuladores no tempo

É extremamente importante, particularmente no estudo das vogais nasais, dadas as suas características dinâmicas, obter informação relativa à variação dos articuladores ao longo do tempo e à forma como se coordenam entre si. Por exemplo, no caso das vogais nasais produzidas em contexto de oclusiva bilabial merece particular interesse os movimentos dos lábios e do véu palatino. A ob-

tenção dessa informação pode ser conseguida utilizando diferentes abordagens. No que concerne ao efetuado no contexto do presente trabalho, a obtenção desses dados proveio dos cortes sagitais, uma vez que ambos os articuladores são visíveis nas imagens obtidas neste plano. Para além disso, poderá vir a ser obtida a partir dos dois cortes coronais marcados na zona do véu palatino e dos lábios.

Para obter essa informação, numa primeira fase deste estudo, recorreu-se a um método proposto por Lammert *et al.* (2010b) e Proctor *et al.* (2011), baseado na variação da intensidade média do *pixel*, através de uma adaptação implementada em MATLAB. Com este método, através das variações de intensidade do *pixel* numa ou em várias regiões de interesse é possível determinar a localização de uma zona de constrição.

A figura 6.8 mostra os diferentes passos necessários para obtenção das curvas de variação dos articuladores ao longo do tempo. No exemplo apresentado, os quatro picos observados (em baixo, à direita) representam o fecho dos lábios (oclusão bilabial) para a produção dos quatro [p] produzidos na sequência. Maior intensidade indica fechamento labial. Se estiverem abertos, a (I_p) reflete IS correspondente ao ar (IS mais baixa nas imagens de RM).

A figura 6.9 apresenta as curvas de variação de intensidade ao longo do tempo para o véu palatino (curva azul) e para os lábios (vermelho) para a sequência de vogal nasal [6~p6, p6~p6, p6~]. O sinal acústico, espectrograma e respetiva anotação fonética são visíveis na parte superior da imagem.

No que diz respeito às vogais nasais, na região dos lábios o método funcionou relativamente bem, como se pode observar na figura 6.9, detetando claramente a oclusão bilabial durante a produção do [p]. Na zona do véu palatino torna-se mais difícil escolher o local exato para colocação da ROI e a dimensão da mesma. Fica-nos a dúvida se o método consegue detetar, com precisão, a altura exata da abertura ou de fecho uma vez que os resultados não foram muito consistentes. Ainda assim, é possível verificar que, na fase final da vogal nasal, os lábios se encontram fechados e o véu palatino está ainda numa posição aberta o que poderá indiciar a emergência de um segmento consonântico nasal. Esta abordagem foi também testada para as séries com consoantes laterais e róticos. Nestes sons os resultados foram melhores podendo vir a ser um método promissor quando essa parte do *corpus* for alvo de análise. Uma descrição mais detalhada do método e alguns dos resultados conseguidos foram publicados em Teixeira *et al.* (2012a).

Posteriormente, foi desenvolvida uma outra abordagem, dentro da nossa equipa, que permitiu obter melhores resultados, em particular, para as vogais nasais, o que é importante considerando os objetivos concretos deste estudo.

Como o RG funcionou relativamente bem nas imagens sagitais permitindo obter os contornos de forma rápida recorreu-se, de novo, a este algoritmo para obter informação um pouco diferente (obtenção de curvas de variação temporal dos articuladores).

De modo semelhante ao que é efetuado com o método baseado na intensidade média do *pixel* é necessário começar por definir as zonas de interesse, neste caso, a nível do véu palatino e dos lábios. Para os lábios, basta que a ROI esteja colocada entre estes. Para o véu palatino,

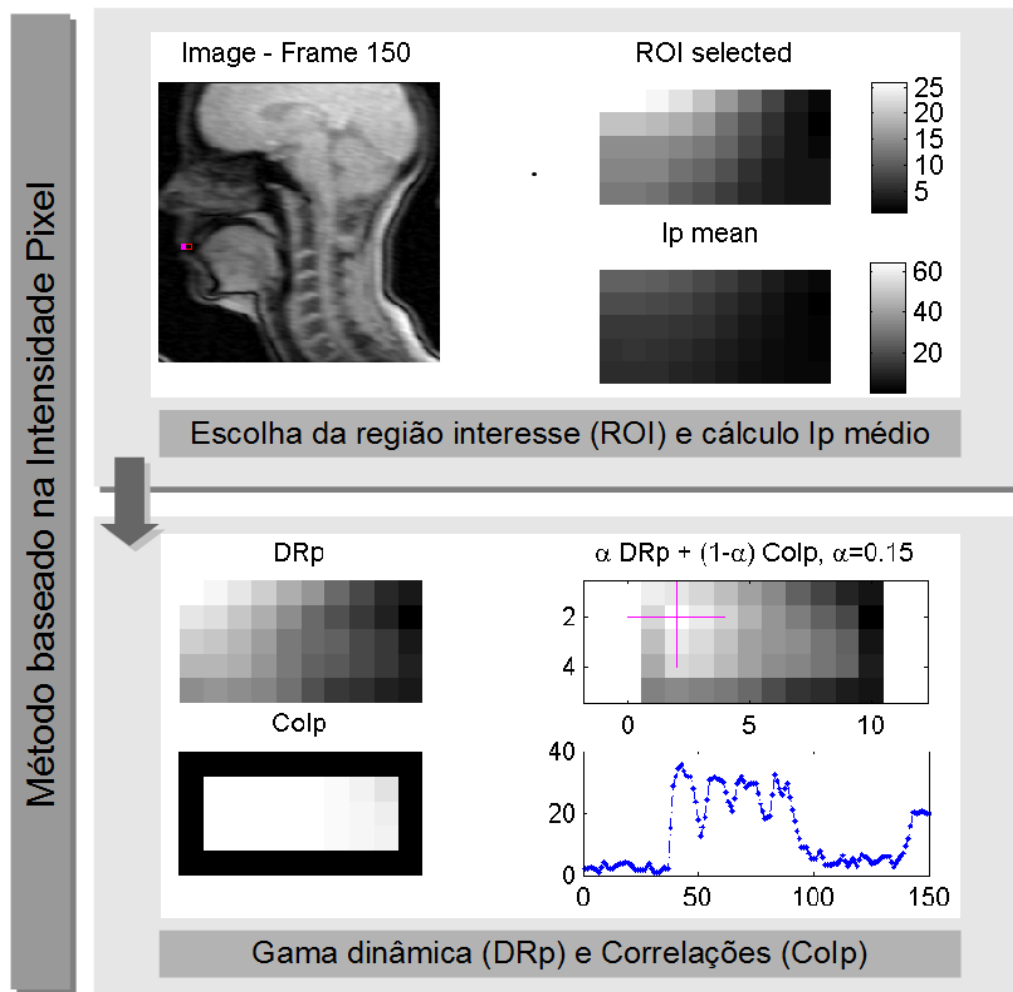


Figura 6.8: Passos referentes à aplicação do método baseado na intensidade média do *pixel* (*Intensity based method*). Imagem adaptada de Teixeira *et al.* (2012a)

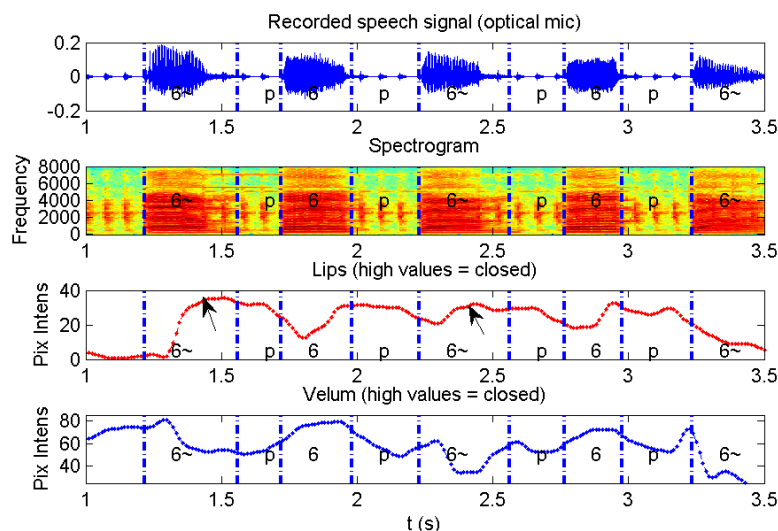


Figura 6.9: Curvas de variação do véu palatino e dos lábios obtidas com o método baseado na intensidade média do *pixel*. Sequência com uma das vogais nasais.

é posicionada, posteriormente, entre a sua face dorsal e a parede posterior da faringe como se mostra na figura 6.10. Para a definição da ROI foi escolhida uma imagem em que o véu palatino está na posição mais baixa e para os lábios uma em que estes se encontrem abertos.

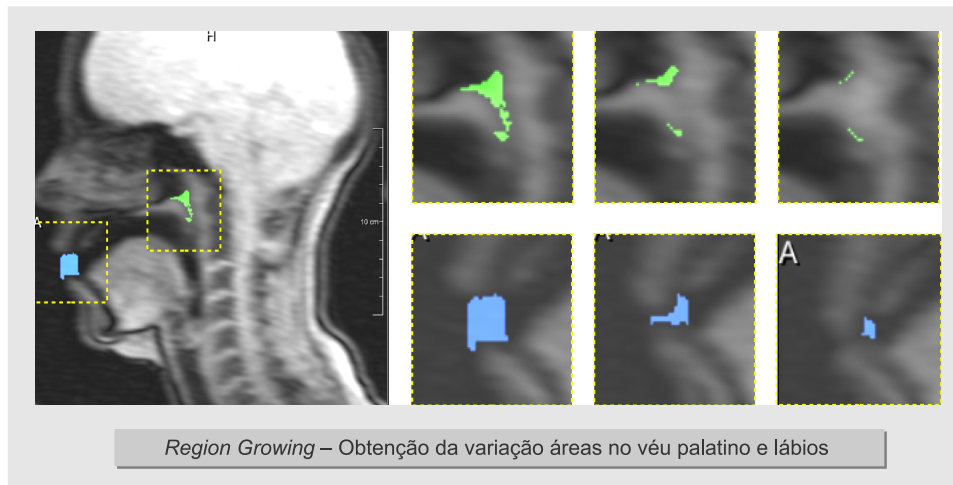


Figura 6.10: Posição do véu palatino (aberto/fechado) e abertura labial obtida a partir dos *pixels* ativos dentro de uma ROI. Imagem publicada em Silva *et al.* (2013a).

Posteriormente, seleciona-se limiar adequado e o RG é aplicado a essa região ao longo de toda a série. O número de *pixels* ativos dentro da ROI é calculado. Quando o véu palatino está baixo, ou os lábios estão abertos o número de *pixels* ativos dentro de cada ROI é maior. Desta forma é possível obter o valor da área nessa região. Note-se que, nesta situação, o valor da área, em si, não tem relevância. O que importa considerar é que, por exemplo, quando a área é pequena os lábios ou o véu palatino estão fechados. O resultado pode ser apresentado sobre a forma de curvas que, através da variação dessas áreas (nas ROI selecionadas), permitem inferir o

posicionamento dos articuladores ao longo do tempo. A figura 6.11 mostra um exemplo das curvas obtidas para o véu palatino (verde - tracejado) e para os lábios (azul) permitindo constatar quando cada um destes está aberto ou fechado. Os resultados obtidos com este método são reportados, mais adiante, na secção 6.5.4. O método descrito está publicado em Silva *et al.* (2013a).

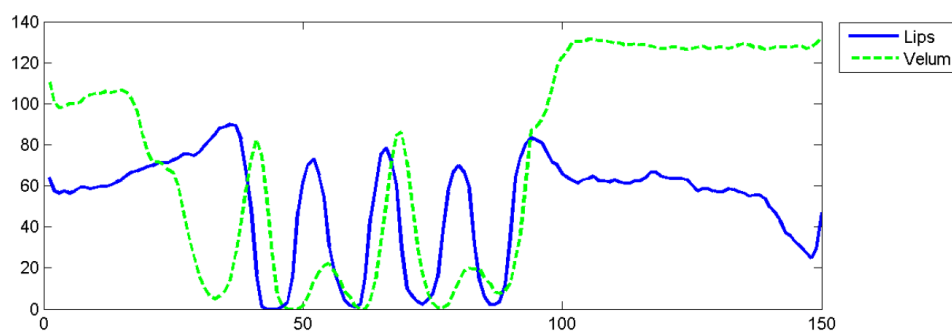


Figura 6.11: Exemplo das curvas obtidas que permitem inferir os movimentos do véu palatino e dos lábios.

6.4.5.3 Obtenção de áreas (nasais/orais) na zona de acoplamento

Os cortes coronais oblíquos obtidos na zona do véu palatino abrem uma oportunidade de obter de forma direta a área do PVF, mas também a área da cavidade oral na zona de acoplamento o que é relevante para o estudo da nasalidade.

Permitem também observar, numa perspetiva diferente das habituais imagens sagitais, os movimentos do véu palatino, ao longo do tempo. A figura 6.12 mostra a orientação dos cortes coronais oblíquos e a identificação das cavidades nasal e oral.

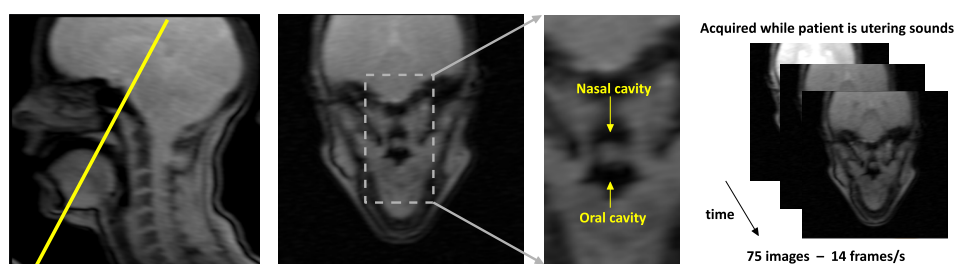


Figura 6.12: Orientação dos cortes coronais no véu palatino (à esquerda), identificação das cavidades nasal e oral (centro) e a série de imagens obtidas ao longo do tempo (direita).

Numa fase inicial, após a aquisição de dados para a primeira falante, recorreremos ao *live-wire* para efetuar a segmentação dos cortes oblíquos utilizando uma das redes *MeVisLab* já utilizada para o processamento das imagens relativas às consoantes laterais. Com este método foram obtidos os primeiros resultados relativos à variação das áreas nasais e orais na zona do acoplamento, publicados em Teixeira *et al.* (2012a). Contudo, apesar de se recorrer a um algoritmo semi-automático, revelou-se muito pouco eficiente o que justificou a procura de outras alternativas que viabilizassem uma extração mais rápida dos dados. Desta forma, aplicou-se o método de RG que permitiu obter melhores resultados, em termos de eficiência, que o *live-wire*.

A figura 6.13 mostra a interface gráfica do *MeVisLab* implementada para o efeito (Silva *et al.*, 2012). O processamento dos dados no que diz respeito à extração de informação proveniente dos cortes oblíquos, marcados na zona do véu palatino, foi dividido em três etapas:

1. Segmentação das imagens para identificação das áreas das cavidades orais e nasais ao longo do tempo;
2. Sincronização e anotação do sinal áudio para fazer identificação das frames de interesse, correspondentes a cada som;
3. Extração e análise das áreas e curvas obtidas.

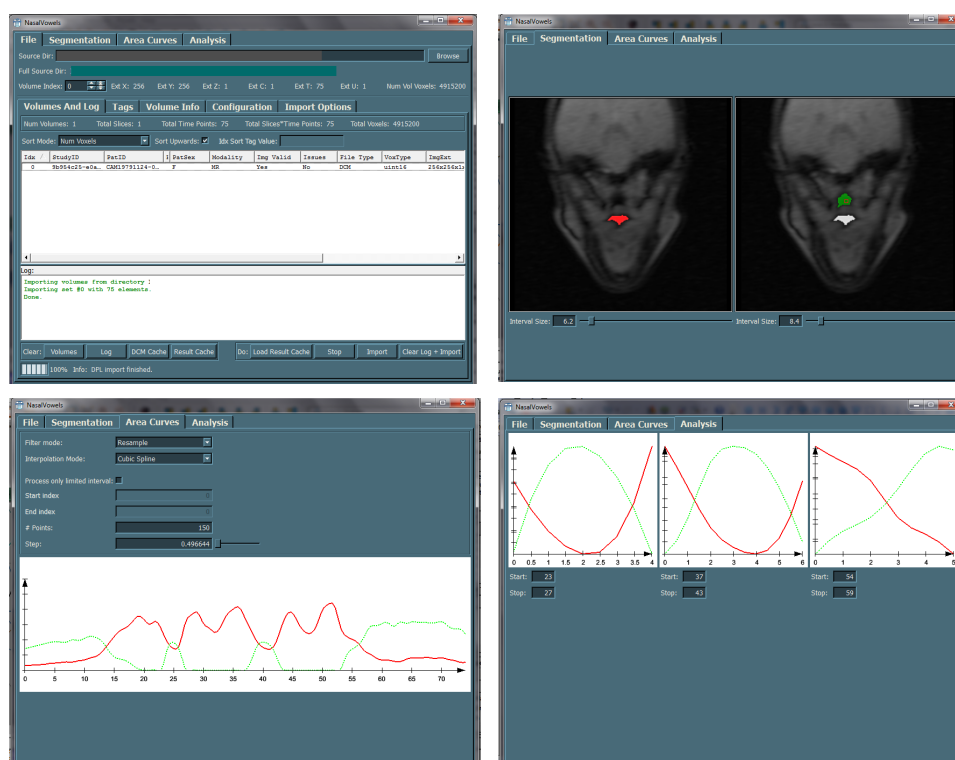


Figura 6.13: Interface gráfica implementada com o MeVisLab, para segmentação das cavidades oral e nasal. Em cima (esquerda), painel para carregar as imagens e à direita a segmentação das cavidades oral e nasal. Em baixo (esquerda), visualização das curvas de área nasal e oral ao longo do tempo; à direita, análise do sub-conjunto de frames relativas a um determinado som. Imagem publicada em Silva *et al.* (2012).

Para efetuar a segmentação com o RG, foi definida uma vizinhança para todas as sementes incluindo o tempo e as coordenadas espaciais da imagem. Desta forma, foi necessário um pequeno número de sementes para efetuar a segmentação de ambas as cavidades. O processo de segmentação começou com a segmentação da cavidade oral, sendo necessário a colocação de uma semente dentro da área de interesse. O intervalo de intensidade, definido automaticamente para o RG, a partir da vizinhança, pode ser ajustado para cada sequência de imagens.

O número de sementes necessárias para segmentar cada cavidade depende do número de vezes que esta fecha durante a sequência de imagens. Como, na maior parte das vezes a

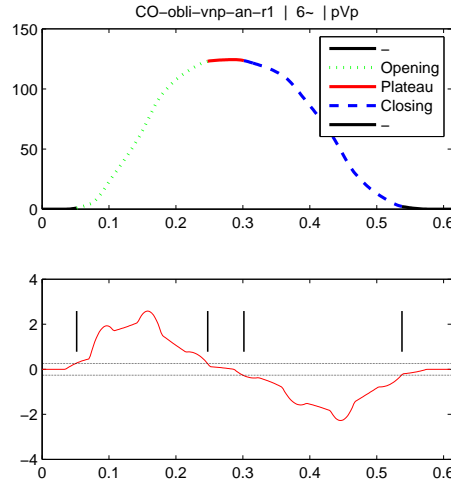


Figura 6.14: Em cima, curva de variação da área nasal com os segmentos correspondentes às diferentes fases assinalados com diferentes cores. Em baixo, primeira derivada da curva de área nasal com as referências utilizadas para determinar a duração do gesto de abertura (*Dopening*), fase estável (*plateau*) e fecho do véu palatino (*Dclosing*).

cavidade oral não fechou, apenas foi necessária a colocação de uma semente. Já para a cavidade nasal, como esta fecha (e.g., quando o véu palatino sobe para a produção da oclusiva), houve necessidade de colocar várias sementes (5 a 6) em várias *frames* ao longo da sequência. O tempo total de segmentação, para cada sequência, rondou os dois minutos sendo necessário, apenas, uma revisão das segmentações, por parte do operador, para efetuar ajustes quando se revelou necessário.

A partir das áreas orais e nasais foi calculada a proporção das áreas nasais e orais ($N/N + O$) e obtida a sua variação temporal.

Com o intuito de obter parâmetros de duração relativos aos movimentos de abertura e fecho do véu palatino recorreu-se a um processo idêntico ao que tem vindo a ser utilizado para o processamento de dados EMA (Teixeira e Vaz, 2001; Oliveira *et al.*, 2009; Nakamura, 2009) e também já utilizado em imagens sagitais de RM em tempo real (veja-se, Byrd *et al.*, 2009). Primeiro, foi efetuada uma sobre-amostragem dos dados (16 vezes) permitindo obter um maior número de pontos nas curvas, tornando-as mais suaves. De seguida, foi calculada a primeira derivada para os segmentos das curvas relativas às áreas nasais. Como referência (*landmarks*) para as transições entre os gestos (e.g., abertura/fecho) foram consideradas as passagens por valores que representam 10% (Oliveira *et al.*, 2009) do valor máximo da primeira derivada (ver figura 6.14).

Como fica patente nas secções anteriores, é necessário encontrar várias alternativas de processamento dos dados para que se consiga tirar partido da vasta e variada informação proporcionada pelas imagens de RM, obtidas com um *corpus* relativamente extenso.

6.5 Resultados

Devido à quantidade de dados adquirida no contexto do presente trabalho, apenas foi explorada parte dos dados, a que se considerou de maior relevância, visando contribuir para a clarificação das questões enumeradas na secção 6.3. Começamos pela descrição e caracterização dos movimentos do véu palatino e da língua (sub-secções 6.5.1 e 6.5.2). Uma vez que são os movimentos destes articuladores que condicionam as áreas nasais e orais na zona de acoplamento, segue-se a apresentação dos resultados obtidos para a variação das áreas (nasais/orais) no tempo e a proporção entre estas (sub-secção 6.5.3). Finalmente, apresentam-se alguns resultados que permitem obter informação acerca da forma como os diferentes articuladores se coordenam entre si, na produção das vogais nasais.

6.5.1 Caracterização dos movimentos do véu palatino

Para descrever os movimentos (abertura/fecho) do véu palatino, no sentido de responder à nossa primeira questão, começamos por analisar as imagens adquiridas sem qualquer tipo de processamento. Através do alinhamento das imagens com o sinal áudio e com a anotação do sinal acústico identificaram-se as *frames* de interesse durante a produção de uma vogal nasal. A figura 6.15 apresenta um conjunto de imagens que incluem os momentos antes, durante e após a produção da vogal [6~] em início de sílaba como em [6~p6].

Como se pode observar na figura 6.15, em posição inicial, o véu palatino encontra-se numa posição de repouso (aberto) antes da falante iniciar a produção da vogal nasal (*fr10*). Começa a subir para fechar (*fr11* a *fr15*) antes da produção da vogal nasal. No início da produção da vogal nasal, o véu palatino está completamente fechado (*fr17*) e assim permanece durante algum tempo. Só depois inicia o gesto de abertura, como é esperado que aconteça na produção de uma vogal nasal (seta amarela). Abre gradualmente (*fr19* a *fr21*) até atingir o máximo de abertura (seta verde). Posteriormente inverte o movimento e começa a subir para a produção do [p] a seguir à vogal. O véu palatino atinge a posição mais elevada na *frame* 23 (seta vermelha, na linha de baixo) que corresponde à fase da oclusão do [p], uma vez que os lábios estão fechados e o véu palatino também está. A *frame* 25 corresponde já à parte inicial da vogal [6], após a explosão da oclusiva (lábios abertos).

Os movimentos do véu palatino durante a produção da vogal nasal em posição medial, como em [p6~p6], são semelhantes (veja-se a figura 6.16). Nesta situação, o véu palatino já se encontra fechado durante a produção do [p] antes da vogal (*fr29* e *fr30*) e assim permanece na fase inicial da vogal nasal (*fr31*).

Quando a vogal é produzida em posição final absoluta (figura 6.17), na fase final da vogal nasal o véu palatino mantém-se aberto (*fr48*), regressando gradualmente à posição de repouso (*fr50*).

Após processamento das imagens sagitais (veja-se parte final da secção 6.4.5.2) pode obter-se a curva de variação do movimento do véu palatino no tempo durante toda a série adquirida.

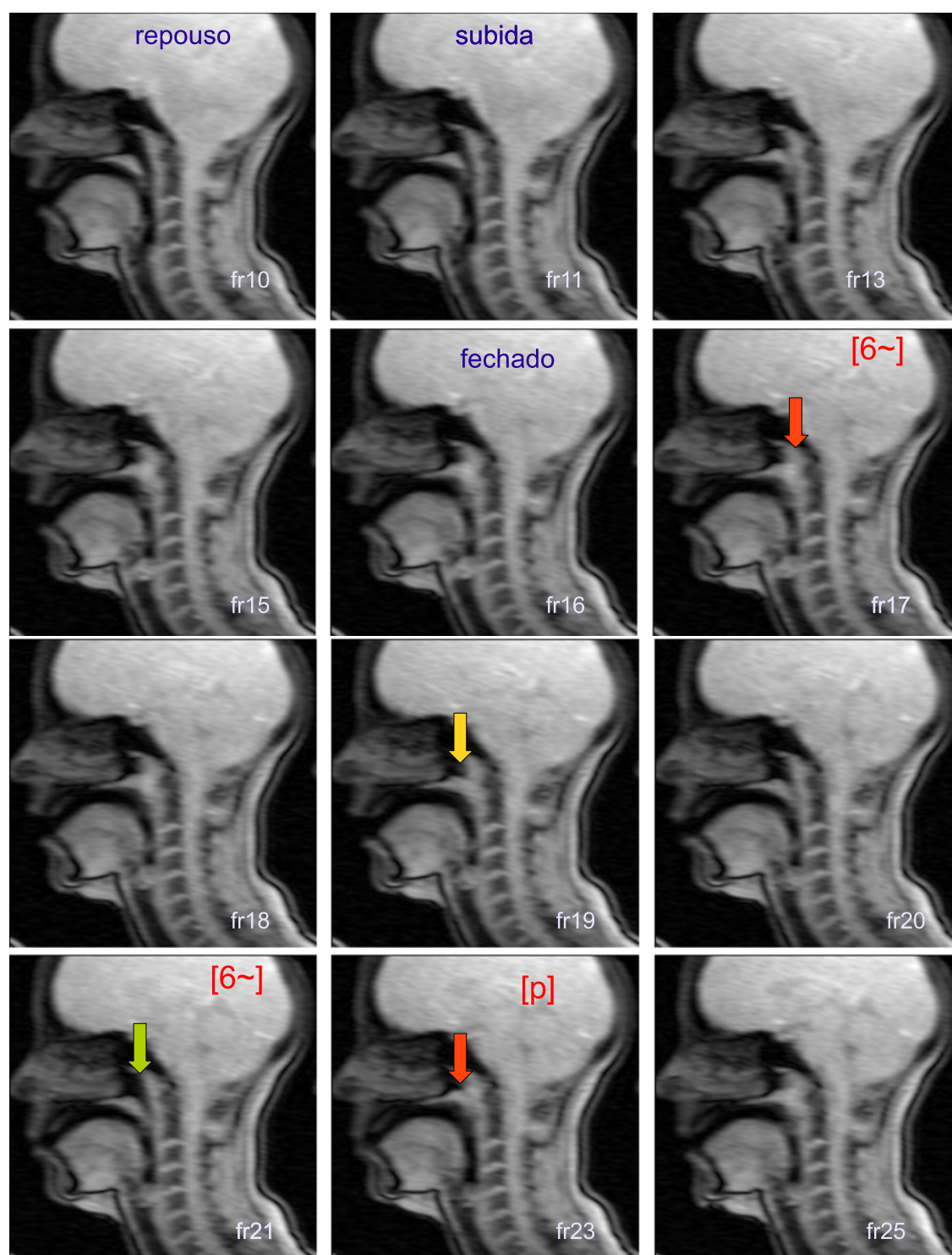


Figura 6.15: Cortes sagitais durante a produção do [6~], em posição inicial, como em **ampa** [6~p6] produzido pela falante CO.

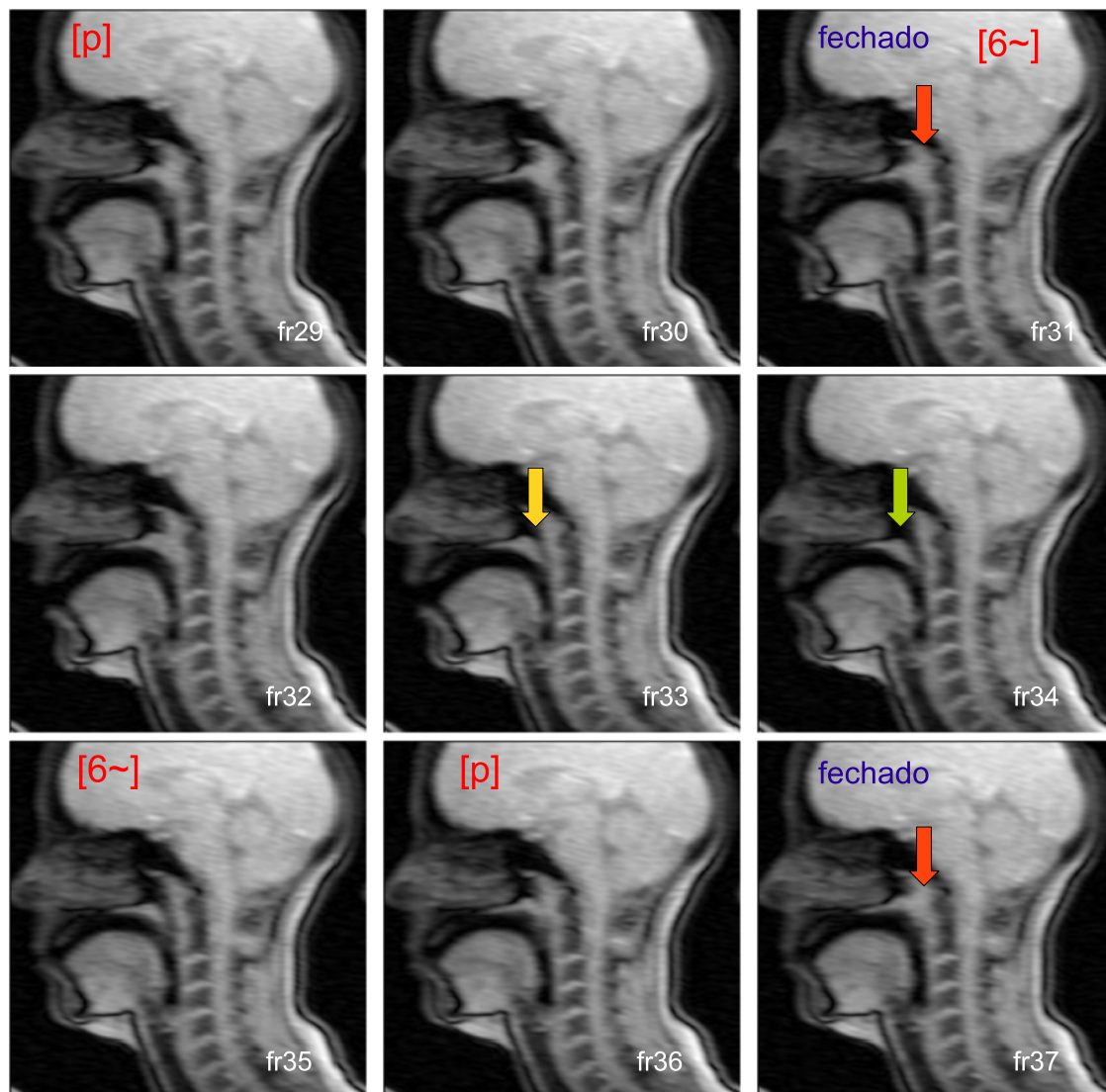


Figura 6.16: Cortes sagitais durante a produção do [6~], em posição medial, como em **pampa** [p6~p6] produzido pela falante CO.

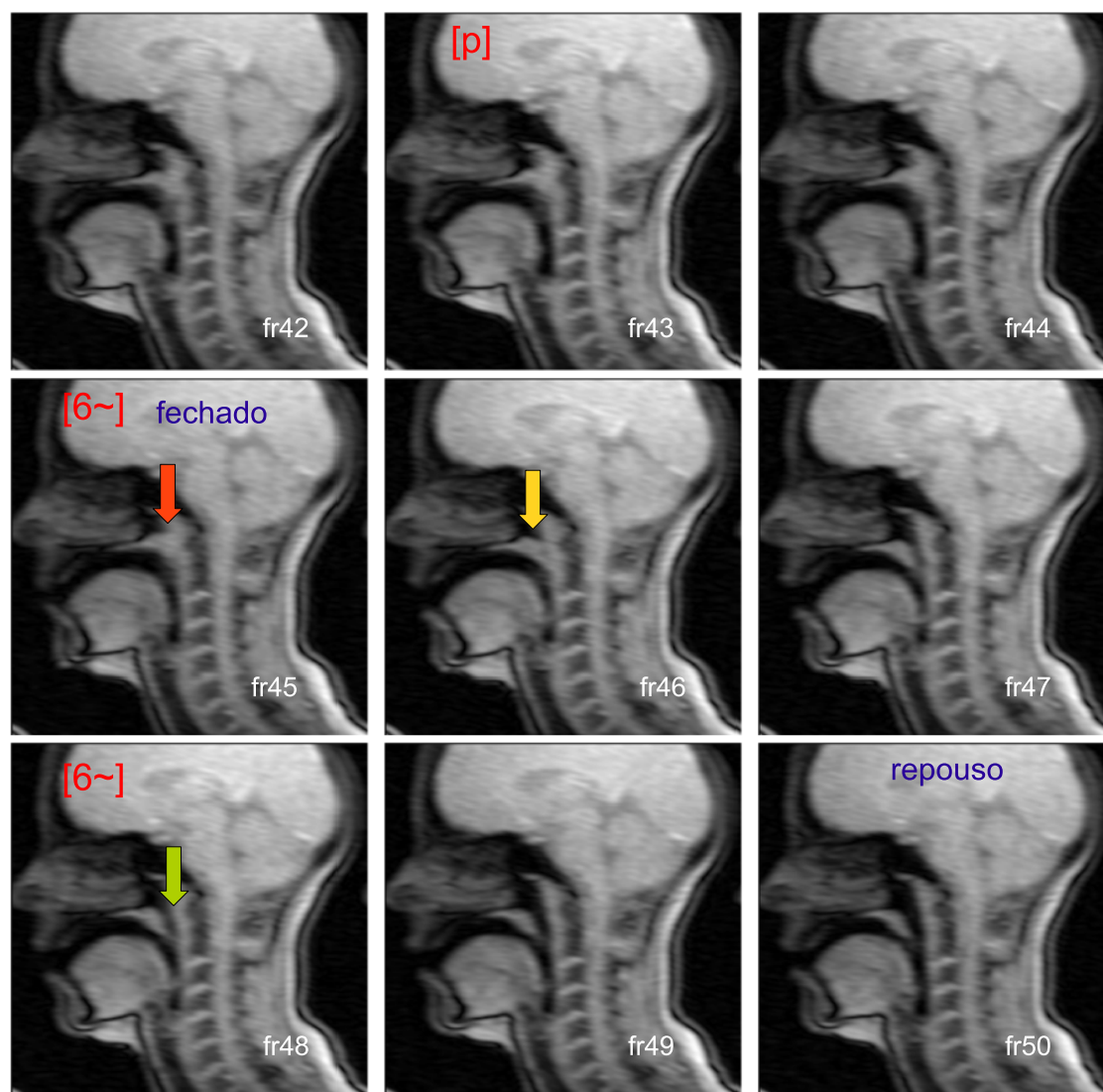


Figura 6.17: Cortes sagitais durante a produção do [6~], em posição final, como em **pan** [p6~] produzido pela falante CO.

A figura 6.18 apresenta um exemplo da falante CO, para as séries com a vogal [6~] (em cima), [i~] (ao centro) e [o~] em baixo. Valor nulo ou baixo na ordenada corresponde a uma posição fechada (ou quase fechada) do véu palatino e o ponto de maior valor de ordenada representa a máxima abertura. Nas imagens apresentadas é possível observar alguns aspetos que descrevemos através das imagens, nomeadamente o facto do véu palatino se encontrar fechado no início da vogal nasal, atingir a sua máxima abertura na fase final da vogal e manter-se aberto quando a vogal é produzida em posição final. Note-se que, mesmo durante a produção da vogal oral [6], o véu palatino não está completamente fechado. O padrão de variação é semelhante, independentemente da vogal nasal considerada.

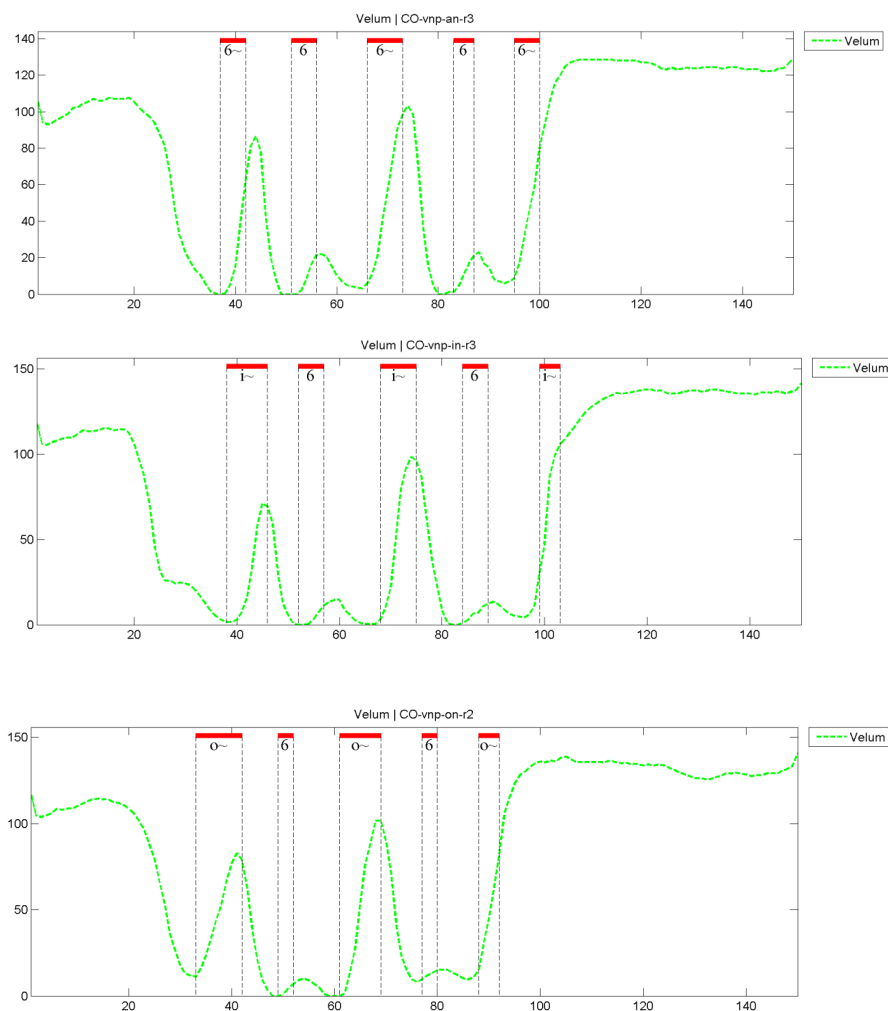


Figura 6.18: Exemplo de curvas que permitem acompanhar os movimentos do véu palatino durante a sequência com as vogais nasais em diferentes posições. Em cima, sequência com a vogal [6~], ao centro [i~] e [o~] em baixo.

6.5.1.1 Duração dos movimentos abertura-fecho do véu palatino

A caracterização dos movimentos do véu palatino passa também pela obtenção dos valores de duração da fase de abertura e fecho do mesmo. É importante demonstrar até que ponto a RM em tempo real com a resolução temporal conseguida (14 *frames/s*) é adequada para a obtenção desses valores. Haveria mais do que uma forma de obter estes resultados. No limite, através da contagem do número de *frames*, seria possível obter uma aproximação dos valores de duração. Todavia, neste estudo, estes valores foram extraídos a partir da segmentação dos cortes coronais obtidos no véu palatino e do processamento efetuado sobre as curvas de variação das áreas na zona de acoplamento, tal como descrito no final da sub-secção 6.4.5.3.

Os resultados obtidos relativos à duração dos movimentos do véu palatino (em *ms*) são apresentados na tabela 6.3 e dizem respeito a duas falantes (CO e CM). Os valores da tabela referem-se ao valores médios ($n = 10$), para a fase de abertura, fase estável, fecho e duração total, para todas as vogais nasais em posição inicial ($M_{inicial}$) e medial (M_{medial}), para duas repetições. No gráfico 6.19 mostra-se apenas a duração média (M_{i+m}) relativa às duas posições, para cada uma das fases e por falante.

Tabela 6.3: *Valores médios da duração (em ms) dos movimentos do véu palatino (abertura, fase estável, fecho e duração total), para as falantes CO e CM, em posição inicial ($n = 10$) e medial ($n = 10$).*

Falante	Posição	Abertura	Estável	Fecho	$DurTot_{A/F}$
CO	$(M_{inicial})$	169.4	38.0	168.8	376.2
	(M_{medial})	214.9	55.1	165.4	435.4
	M_{i+m}	192.2	46.6	167.1	405.8
CM	$(M_{inicial})$	178.7	14.9	125.6	319.1
	(M_{medial})	149.0	40.5	130.5	320.0
	M_{i+m}	163.8	27.7	128.0	319.5

Como se pode observar através da análise da tabela, a duração total do ciclo de abertura e fecho do véu palatino, para as duas posições é de 405.8 *ms* para a falante CO e 319.5 *ms* para a falante CM. Esta diferença de valores, entre as falantes, reflete-se também nos valores de cada uma das fases sendo sempre mais elevados na falante CO.

A duração do ciclo de abertura/fecho do véu palatino, para cada uma das posições consideradas (inicial e medial), é muito semelhante no caso da falante CM, enquanto que para a falante CO o valor da duração total do ciclo de abertura/fecho do véu palatino é mais elevado em posição medial (435.4 *ms*) do que em posição inicial (376.2 *ms*).

O valor médio ($n = 20$) da duração, para cada uma das fases (fase de abertura, fase estável e fecho), considerando as duas posições foi, respetivamente 192.2/46.6/167.1 *ms* para a falante CO e 163.8/27.7/128.0 *ms* para a falante CM.

Constata-se que a duração do movimento de abertura é superior ao de fecho para ambas as falantes. A diferença é de cerca de 36 *ms* (11% da duração total do ciclo) para a falante CM e 25 *ms* (6% da duração total do ciclo) para a falante CO.

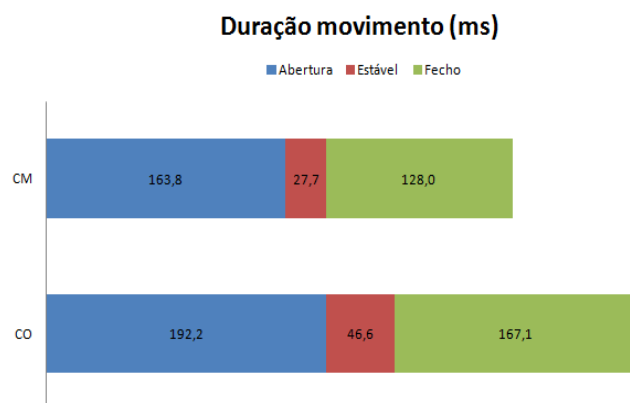


Figura 6.19: Valores médios da duração (em *ms*) dos movimentos de abertura, zona estável e fecho do véu palatino para as falantes CO e CM.

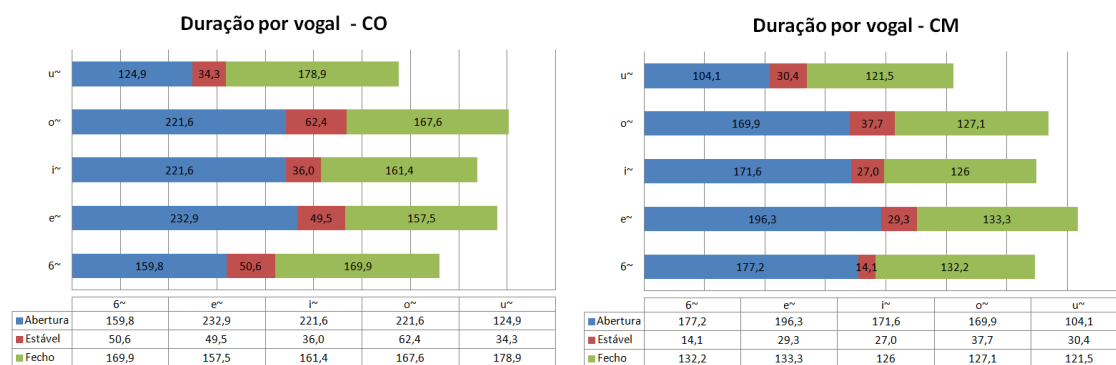


Figura 6.20: Valores médios da duração (em *ms*) dos movimentos de abertura, zona estável e fecho do véu palatino para as falantes CO e CM, para cada uma das cinco vogais nasais do PE [6~],[e~],[i~],[o~] e [u~].

Pese embora não se possa fazer aplicação de testes estatísticos, uma vez que o número de amostras é muito baixo ($n = 4$), tentamos perceber se haveria algum efeito da vogal na duração de cada uma das fases e na duração total do ciclo abertura/fecho do véu palatino. O gráfico 6.20 apresenta os valores médios da duração para cada uma das fases, por falante, em função de cada uma das vogais nasais.

Um resultado relevante e observado para as duas falantes, é que a vogal [u~] contraria a tendência geral observada para todas as outras vogais nasais, apresentando um valor de duração do movimento de abertura mais curto que o de fecho. De todas as nasais é o [u~] que apresenta uma abertura mais rápida. O valor mais longo de abertura foi registado para a vogal [e~]. A duração total do ciclo de abertura e fecho do véu palatino foi maior para as vogais [e~] e [o~].

6.5.2 Articulação oral das vogais nasais

Para além do movimento do véu palatino, é importante analisar os ajustes articulatórios a nível dos articuladores orais (e.g., língua ou lábios) durante a produção das vogais nasais. No que diz

respeito à língua, por exemplo, uma mera observação das imagens poderá não permitir detetar a existência de variação ao longo da produção da vogal. Neste caso, revelou-se fundamental a extração e sobreposição de contornos para detetar diferenças. Os contornos foram obtidos a partir das imagens sagitais, com recurso ao método descrito na secção 6.4.5.1.

Para estabelecer a comparação entre as vogais nasais e as congéneres orais foram efetuadas comparações dos contornos para os pares orais/nasais, para todas as falantes e vogais nasais. Para a vogal oral foi escolhido o contorno correspondente a uma *frame* a meio da produção da vogal. No caso da vogal nasal, a *frame* selecionada para efetuar a comparação foi aquela em que se visualiza o véu palatino na posição mais descida (máxima amplitude) e com os lábios ainda abertos (o que poderá não corresponder exatamente ao meio da vogal). Esta escolha tem em consideração resultados prévios, obtidos para o PE, em que se verifica que o véu palatino atinge o seu *target* na parte final da vogal nasal (cf. Oliveira, 2009, p. 240).

Nesta secção, apresentam-se os resultados obtidos através da sobreposição e comparação de contornos dos pares vogal oral/nasal: [i]/[i~] e [u]/[u~]. A nasal [6~] é comparada com as orais [6] e [a], o [e~] com [E] e [e], o [o~] com [O] e [o]. O contexto considerado para as comparações foi de vogal nasal em posição medial em contexto de oclusiva bilabial [p], uma vez que as vogais orais foram adquiridas num contexto similar (e.g., [p6~p6] *versus* [pap6] e [p6p6]).

A figura 6.21 apresenta a comparação para os pares [i]/[i~] e [u]/[u~] para as falantes CO e SV. Como se pode observar, para além da posição do véu palatino, que está mais baixo na produção da vogal nasal, os ajustes ao nível dos outros articuladores são insignificantes. No [u~], aqui exemplificado através da falante CO, há um pequeno avanço da ponta da língua comparativamente à vogal oral, que também se verifica para a falante CM. Para a falante SV, como se pode constatar na imagem, acontece precisamente o oposto.

A vogal [6~] foi comparada com a vogal oral mais aberta [a] e com o [6] (ver figura 6.22). Verificou-se que, para qualquer uma das falantes, a configuração articulatória do [6~] (contorno vermelho) é mais próxima da do [6] (azul) do que do [a] (verde), confirmando a tendência já observada em estudos anteriores para o PE (e.g., Martins *et al.*, 2007). Considerando o par [6]/[6~], à semelhança do que acontece com as vogais [i~] e [u~], os ajustes articulatórios revelaram-se também muito subtis, sendo os contornos praticamente sobreponíveis nas falantes CO e SV e existindo uma discreta subida do dorso para a falante CM.

As vogais em se observou a existência de maiores ajustes articulatórios, foram as vogais [e~] e [o~]. Considerando o [e~], no caso da falante CO, a nasal é claramente produzida com uma configuração da língua muito mais próxima do [E] do que do [e]. No entanto, existe na nasal uma menor abertura labial do que no [E] e [e]. Contudo, para as outras duas falantes a vogal nasal parece ser produzida através de uma configuração intermédia entre o [E] e o [e] e não existem diferenças na abertura dos lábios.

A sobreposição dos contornos relativos às comparações do [o~] com as vogais orais é apresentada na figura 6.22 (em baixo). Do conjunto, das vogais, esta foi a que apresentou maiores ajustes ao nível do dorso da língua, sobretudo anteriormente, na zona da ponta da língua.

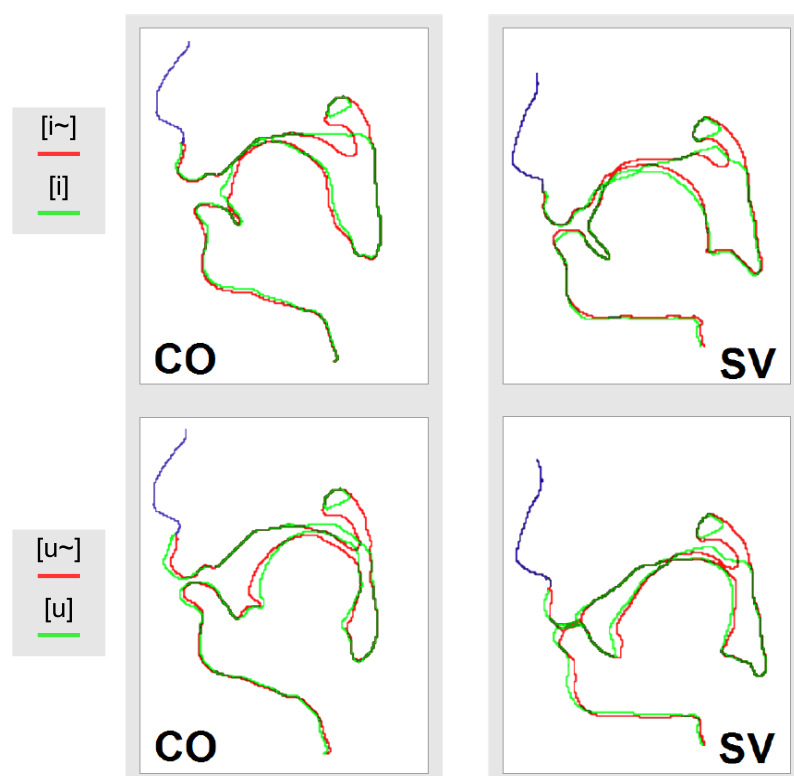


Figura 6.21: Comparação da articulação oral/nasal efetuada através da sobreposição de contornos relativos aos pares [i]/[i~] (em cima) e [u]/[u~] (em baixo) para as falantes CO (esquerda) e SV (direita). Imagem publicada em Oliveira *et al.* (2012).

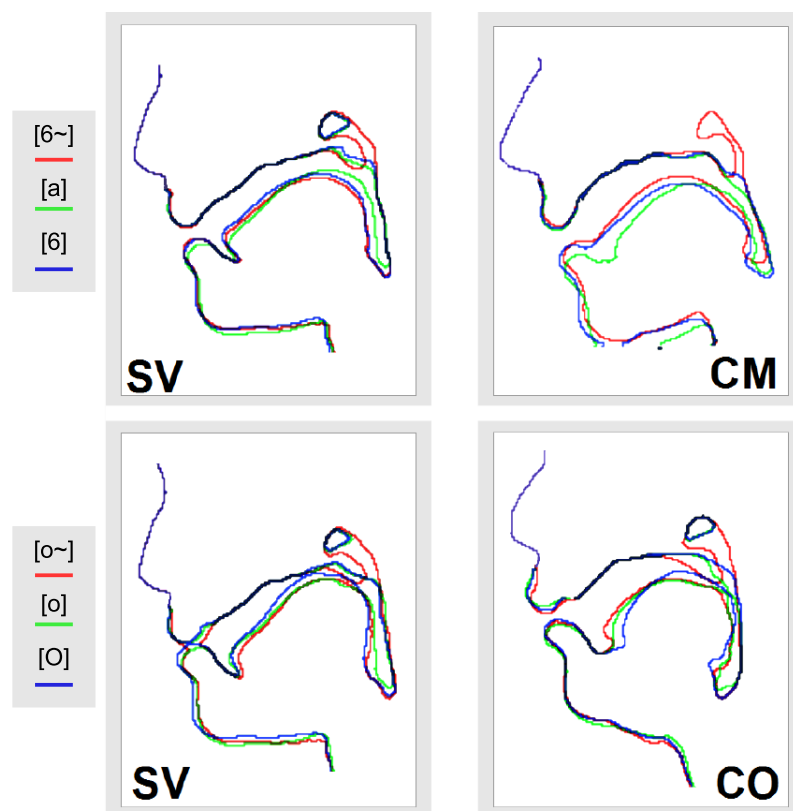


Figura 6.22: Comparação da articulação oral/nasal através da sobreposição dos contornos relativos às vogais [a]/[6]/[6~] (em cima) e [O]/[o]/[o~] (em baixo) para as falantes SV (esquerda) e CM (direita). Imagem publicada em Oliveira *et al.* (2012).

Para além das comparações efetuadas, é ainda possível observar a variação da língua e dos outros articuladores para cada uma das vogais nasais, ao longo do tempo, através da sobreposição de três *frames* obtidas em diferentes momentos da produção da vogal nasal.

As imagens das figuras 6.23 e 6.24 dizem respeito à falante CO. A figura 6.23 mostra as produções de [6~] (em cima), [e~] (centro) e [i~] (baixo). Na figura 6.24 apresenta-se informação similar, mas para as vogais [o~] (em cima) e [u~] (em baixo).

A análise da figura permite-nos verificar que também são as vogais [i~] e [u~] que apresentam menor variação ao longo do tempo, mantendo-se o dorso da língua sensivelmente na mesma posição com abaixamento progressivo do véu palatino ao longo da vogal. As vogais que sofrem mais ajustes articulatórios orais são as vogais [e~] e [o~], particularmente a nível do dorso da língua.

No caso do [e~], há um recuo do dorso da língua na fase final da produção da vogal nasal condicionando, neste caso, uma zona faríngea mais constrangida. No [6~], para esta falante, não há ajustes dignos de referência a nível do dorso da língua, mas há um fechamento dos lábios na fase final da produção da vogal (traçado a verde). Como se constata, quando há algum tipo de ajuste, ele verifica-se nas três posições silábicas consideradas.

No *corpus* foram ainda incluídas algumas produções em que a vogal nasal foi produzida de forma isolada para verificar a existência, ou não, de diferenças a nível da posição do véu palatino ou de outros articuladores (e.g., língua) em relação às vogais produzidas em contexto. Neste caso em particular, foi efetuada comparação com a vogal em posição medial no contexto de [p].

A figura 6.25 apresenta um exemplo dessa comparação para a falante CM. Podemos verificar que nas vogais produzidas de forma isolada, para além do movimento de descida do véu palatino há uma grande variação na posição da língua ao longo da produção da vogal que se reflete também, quando a língua recua, a nível das dimensões da faringe. Existem, também, modificações a nível da posição da mandíbula e na abertura dos lábios, particularmente, do lábio inferior. Relativamente à posição do véu palatino, não se verificou existirem diferenças dignas de registo entre as produções produzidas isoladamente e em contexto.

Note-se, no entanto, que estas alterações dos articuladores, ao longo do tempo, nas vogais isoladas são muito mais marcadas nas vogais [e~] e [o~] do que para vogais [i~] e [u~] (cf. figura 6.26) para produções das falantes CO e SV.

De realçar também que na última *frame* relativa à vogal nasal, produzida em contexto (com oclusiva bilabial), se verifica que os lábios já se encontram fechados e que o véu palatino permanece ainda aberto. Este facto é compatível com a emergência de uma consoante nasal homorgânica com a consoante seguinte (neste caso um [m]), indo ao encontro do que é reportado para o PE por vários autores (Oliveira, 2009; Domingos, 2011; Ashby *et al.*, 2011). Como seria de esperar, essa oclusão labial não acontece quando a vogal é produzida de forma isolada, mas há nesta situação um estreitamento no trato oral, cuja localização varia com a vogal, na fase final da sua produção.

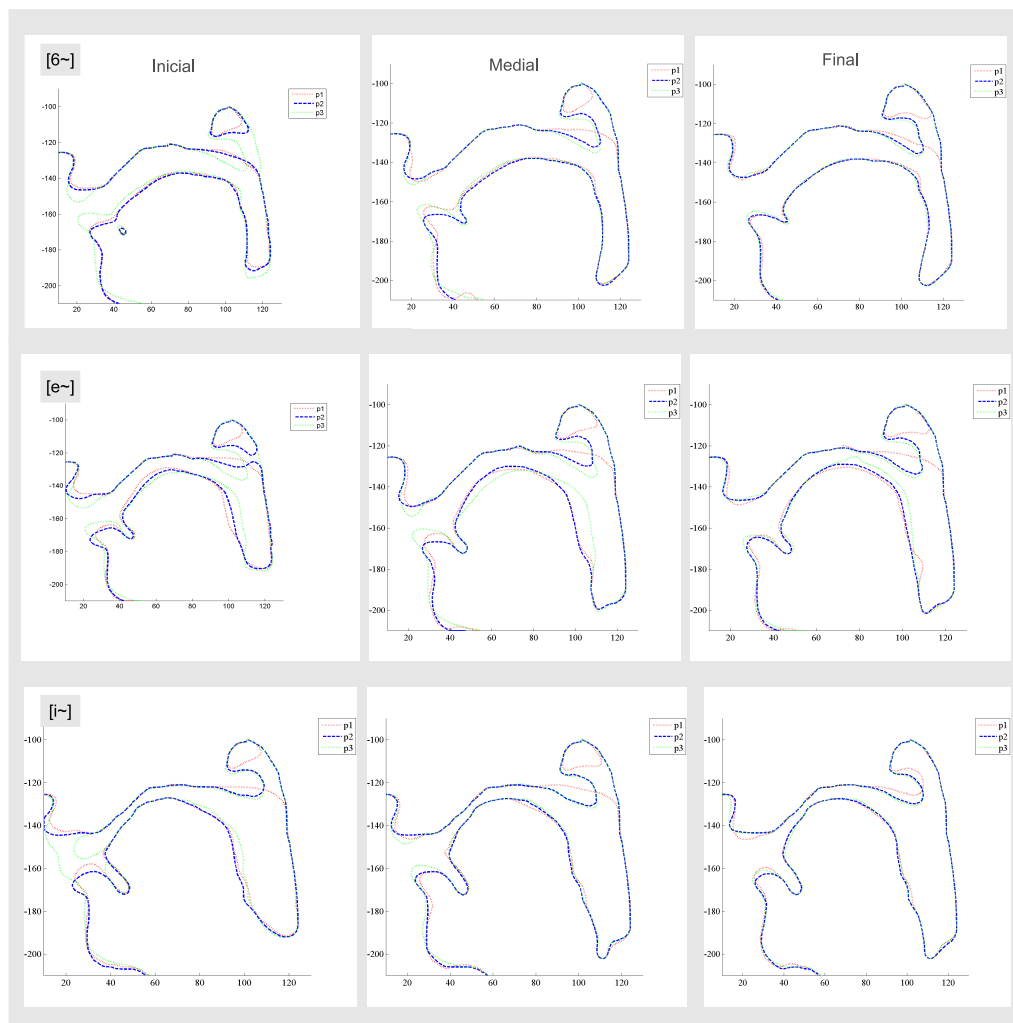


Figura 6.23: Sobreposição de *frames* obtidas em diferentes momentos temporais ao longo da produção das vogais nasais [6~] (em cima), [e~] (ao centro) e [i~] (em baixo) produzidas pela falante CO. Da esquerda para a direita - produção de cada vogal em posição inicial, medial e final.

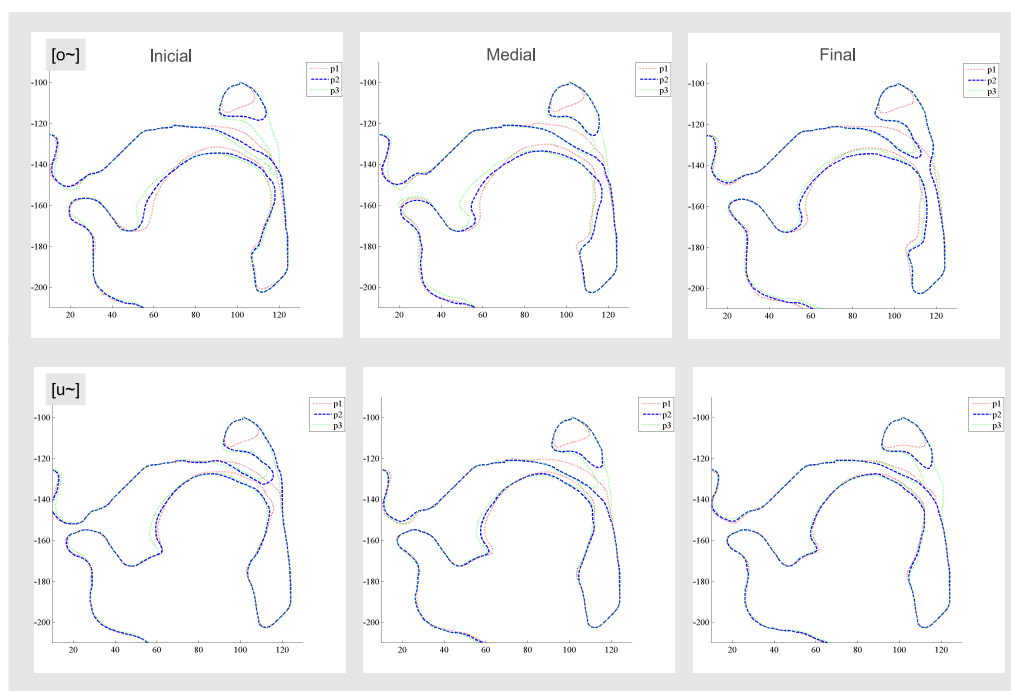


Figura 6.24: Sobreposição de *frames* obtidas em diferentes momentos temporais ao longo da produção das vogais nasais [o~] (em cima), e [u~] (em baixo) produzidas pela falante CO. Da esquerda para a direita - produção de cada vogal em posição inicial, medial e final.

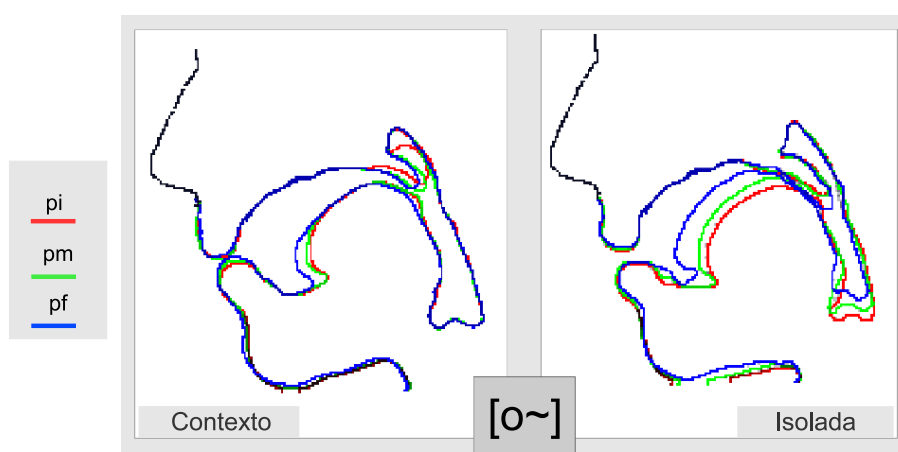


Figura 6.25: Sobreposição dos contornos obtidos ao longo do tempo (*frames* relativas ao início, meio e final) durante a produção da vogal [o~] produzida de forma isolada (esquerda) e em posição medial em contexto de [p] como em [po~p6]. Exemplo ilustrativo para a falante CM.

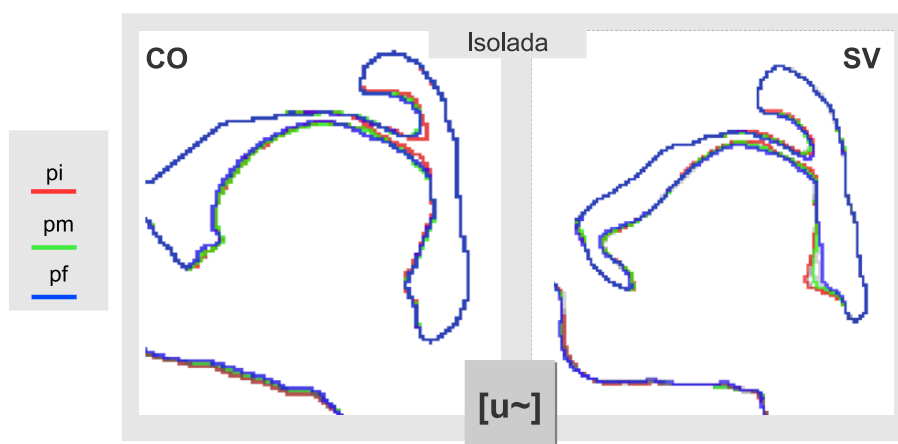


Figura 6.26: Sobreposição dos contornos obtidos ao longo do tempo (*frames* relativas ao início, meio e final) durante a articulação da vogal [u~] produzida de forma isolada pela falante CO (esquerda) e pela falante SV (direita).

6.5.3 Áreas nasais e orais na zona de acoplamento

O movimento do véu palatino, e mesmo de outros articuladores orais, afeta as áreas na zona de acoplamento do trato oral e nasal. Os resultados apresentados nesta secção resultam da informação recolhida a partir do corte coronal oblíquo adquirido na zona do véu palatino, ao longo do tempo.

6.5.3.1 Variação temporal

Após processamento das imagens coronais, obtiveram-se os valores das áreas nasais e orais, na zona de acoplamento e, a partir destas, as curvas de variação ao longo do tempo. Os dados resultam das produções das três falantes, todas as vogais nasais e posições silábicas, em contexto de oclusiva bilabial [p]. Na figura 6.27, pode observar-se um exemplo representativo (referente à falante CO), das curvas de variação das áreas nasais/orais ao longo do tempo. A curva da área oral é representada a vermelho (tracejado) e a área nasal corresponde a curva azul. Tendo por base a anotação fonética efetuada, foi considerado o período correspondente à vogal nasal e também o [p] a seguir à vogal.

Como se pode constatar pela análise da figura, no início da vogal nasal as áreas nasais são muito pequenas (abaixo dos 25 mm^2), denotando uma pequena abertura do PVF. Depois, observa-se um aumento gradual dessas áreas até ser atingido o valor máximo de área nasal que corresponderá à altura em que o véu palatino está na sua posição mais baixa. Finalmente, as áreas começam a decrescer à medida que o véu palatino sobe para a produção da oclusiva [p]. A magnitude das áreas nasais é bastante similar para as diferentes vogais nasais (aproximadamente 100 mm^2 , para esta falante). Este é um padrão comum para as vogais nasais produzidas em posição inicial (e.g., [6~p6] e medial [p6~p6]). Em posição final, depois de atingir a área nasal máxima, esta mantém-se constante, o que pressupõe que o véu palatino ficará aberto, em posição de repouso.

As áreas da cavidade oral, na zona de acoplamento, seguem uma tendência inversa.

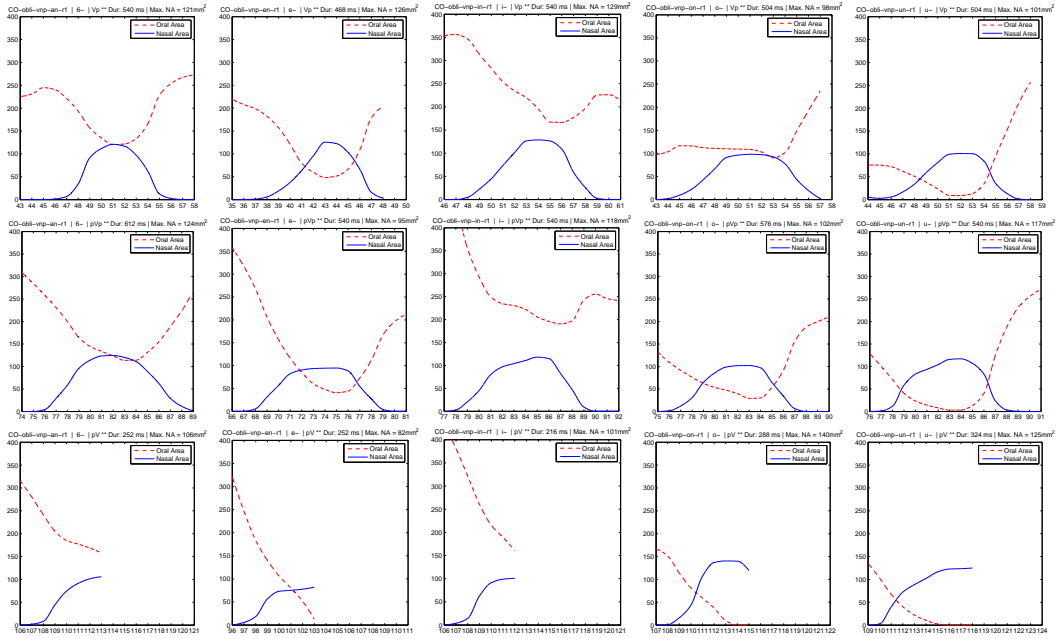


Figura 6.27: Exemplo da variação das áreas orais e nasais ao longo do tempo (falante CO) para todas as vogais nasais (da esquerda para a direita - [6~], [e~], [i~], [o~], [u~]) produzidas em início de palavra (em cima), posição medial (ao centro) e posição final (em baixo).

Começam por diminuir, atingem um valor mínimo e aumentam de novo na fase final. As áreas orais, contrariamente ao que se verifica para as áreas nasais, variam bastante em função das diferentes vogais nasais. Esta diferença na magnitude das áreas orais deve-se à posição do dorso da língua na cavidade oral que, naturalmente, será diferente em função da altura da vogal nasal e condicionará a dimensão da passagem oral. O dorso da língua está mais elevado na produção de uma vogal alta e recuada como o [u~] e mais baixo na produção do [6~]. Note-se que para a vogal [u~], no ponto temporal em que a área nasal é máxima, a área oral é praticamente nula. Já para o [6~] quando a área nasal é máxima a área oral é superior a 100 mm², ou seja, similar à área nasal. Para o [i~], a área oral é sempre superior à área nasal. Neste caso, a língua está numa posição elevada mas muito avançada permitindo áreas maiores da passagem oral, na zona do acoplamento, mesmo com o véu palatino numa posição baixa. De um modo geral, as falantes CM e SV seguem uma tendência semelhante.

6.5.3.2 Área máxima de abertura do PVF

A área máxima de abertura do PVF corresponderá à posição em que o véu palatino está numa posição mais baixa. Na tabela 6.4 apresentam-se os valores da média ($n = 30$), mediana, máximo, mínimo e DP relativos à área máxima de abertura do PVF, por falante, para as produções efetuadas (para duas repetições de cada série). A análise da tabela permite-nos constatar que existem diferenças, entre as falantes, na área máxima de abertura do PVF. A falante CO é aquela em que a área máxima de abertura do PVF é, em média, mais elevada. As áreas mais pequenas foram observadas para a falante CM. Esta falante é, no entanto, aquela em que há menor variabilidade.

Esta diferença na magnitude das áreas entre as falantes, apesar de não muito acentuada, reflete as diferenças anatômicas existentes entre as falantes.

Falante	Média ($n = 30$)	Mediana	Máximo	Mínimo	DP
CO	109.8	114.0	140.0	77.0	17.0
SV	93.7	93.5	125.0	60.0	18.3
CM	82.7	82.0	104.0	64.0	9.1

Tabela 6.4: Valor médio, mediana, máximo, mínimo e desvio padrão para os valores de máxima abertura do PVF (mm^2), por falante.

A área nasal máxima (valor médio relativo a seis produções), para cada uma das vogais nasais, por falante, apresenta-se no gráfico da figura 6.28 (à esquerda). A sua análise permite-nos constatar que para a falante CO os valores variam entre os $102.2 mm^2$ para o [6~] e $116,8 mm^2$ para o [o~], na falante SV o valor mais baixo verificou-se para o [e~] com $84.1 mm^2$ e o mais elevado para as vogais [6~] e [u~] com $98.5 mm^2$. Para a falante CM o valor mais baixo é de $79.3 mm^2$ para a vogal [6~] e $86,8 mm^2$ para a vogal [u~].

Tendo em atenção as várias vogais nasais não se observou um padrão de variação consistente para as três falantes, mas as áreas do PVF não divergem muito entre as vogais. Veja-se, por exemplo, a falante CM em que não há praticamente diferença na abertura do PVF entre as diferentes vogais; para as falantes SV e CO a diferença máxima situa-se à volta de $14.5 mm^2$. Estes valores de área do PVF corresponderão a diferenças muito subtis na altura do véu palatino. Finalmente, o gráfico da figura 6.28 (à direita) apresenta as áreas nasais máximas, para as três falantes, em função da posição. Também neste caso as diferenças não são relevantes, pelo menos para as falantes CO e CM. No caso da falante SV, a diferença de abertura entre a posição inicial e medial é um pouco maior.

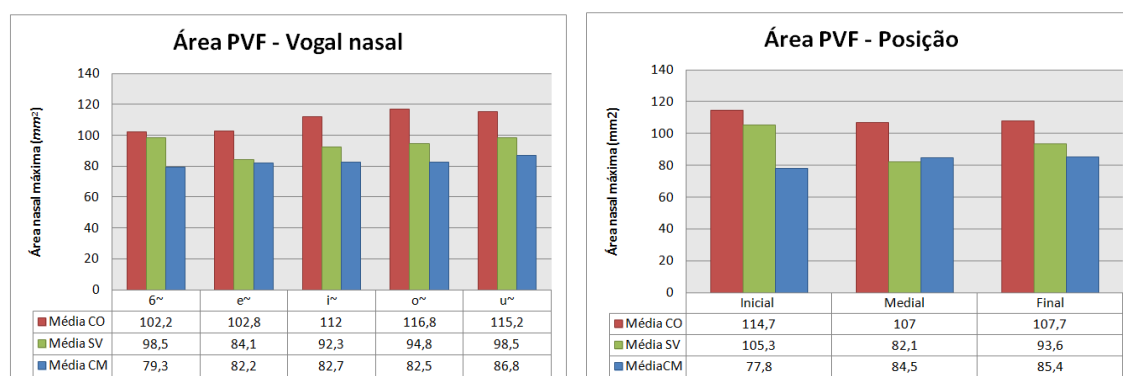


Figura 6.28: Área máxima nasal ou abertura do PVF (mm^2) em função da vogal nasal (à esquerda) e posição silábica (à direita) para as 3 falantes.

6.5.3.3 Proporção nasal/oral (P_{NO})

As áreas e a relação das áreas, na zona do acoplamento, parece-nos de extrema relevância, tendo em consideração as simplificações usuais que são feitas para a nasalidade já referidas no capítulo 4. De facto, os aspetos relacionados com a nasalidade vão muito para além da constatação simples do véu palatino estar aberto ou fechado (o véu palatino pode estar aberto e o som produzido não ser percebido como nasal e podem ser produzidas ressonâncias nasais com o véu palatino fechado).

Muitos estudos se têm debruçado sobre estes assuntos, muitas vezes obtendo informação de forma indireta (e.g., inferindo áreas a partir de pontos ou a partir das distâncias no plano médio-sagital medidas sobre traçados radiográficos). Outros tentaram obter a informação das áreas e sua proporção a partir de processos de inversão, devido à inexistência de dados articulatórios diretos (Feng e Castelli, 1996; Rossato *et al.*, 1998). Neste estudo é possível ter acesso privilegiado à zona do PVF, através dos cortes coronais aí obtidos, e medir essas áreas de forma direta o que nos parece relevante. Desta forma, depois de ter os valores das áreas nasais e orais foi calculada a proporção entre estas utilizando a relação que designamos por $(P_{NO})^{10}$. O facto de se considerar uma proporção, que é uma medida relativa, é vantajoso uma vez que minimiza o efeito da existência de diferenças nas dimensões do trato vocal entre falantes (Engwall *et al.*, 2006). A razão para se considerar, em denominador, a soma da área nasal e oral e não apenas a área oral resulta da necessidade de minimizar o efeito de saturação quando o véu palatino se aproxima muito da língua na produção de vogais nasais posteriores ou recuadas ou em consoantes nasais. No caso de consoantes nasais como a área oral é zero o valor tenderia para infinito (Rossato *et al.*, 1998).

Assim, o (P_{NO}) foi calculado da seguinte forma:

$$P_{NO} = N/(N + O) \quad (6.2)$$

Onde (N) corresponde à área nasal e (O) à área oral, na zona de acoplamento, em cada ponto da sequência temporal. Foi obtida a curva de variação do P_{NO} , ao longo do tempo.

Como se antecipava, a curva de variação do P_{NO} , ao longo do tempo, segue um padrão semelhante ao observado para as áreas nasais. No início de qualquer uma das vogais nasais, é baixo, quase zero, e aumenta ao longo da produção da vogal nasal. Contudo, como para o cálculo da proporção concorrem também as áreas orais e estas são diferentes em função da vogal, também o valor do P_{NO} é diferente para as diferentes vogais nasais.

Foi calculada a média do valor máximo da proporção P_{NO} nas diferentes posições silábicas e repetições efetuadas, para cada falante e vogal nasal. Os resultados são apresentados na tabela 6.5.

Como se pode verificar pela análise da tabela, para todas as falantes, o P_{NO} máximo é mais elevado para as vogais mais recuadas ([o~] e [u~]) do que para as vogais mais anteriores

¹⁰ Apesar de existir alguma semelhança, esta proporção não deve ser interpretada exatamente como o N% proposto por Reenen (1982), uma vez que no cálculo do N% ($N\% = N/(N + MC) \times 100$) o N representa a área nasal na zona do acoplamento (*Nose coupling*), mas o MC (*Mouse coupling*) representa a área de maior constricção ao longo do trato oro-faríngeo que, no caso de uma vogal como o [i], não será na zona do acoplamento, mas segundo os cálculos do autor, cerca de 4 cm mais adiante no trato oral. A proporção que aqui é apresentada é mais semelhante com o *area ratio* proposto por Feng e Castelli (1996) e utilizado também por Rossato *et al.* (1998).

Falante	[6 ~]	[e ~]	[i ~]	[o ~]	[u ~]
CO	0.47	0.66	0.39	0.60	0.97
SV	0.30	0.30	0.35	0.76	0.99
CM	0.39	0.47	0.34	0.96	0.99

Tabela 6.5: Valor médio respeitante ao valor máximo do P_{NO} , por falante e para cada vogal nasal.

([e~] e [i~]) ou para a vogal [6~] que é mais baixa. No caso do [u~] os valores são muito próximos de um o que implica uma abertura oral muito pequena.

6.5.4 Organização temporal dos movimentos dos articuladores durante a produção de vogais nasais

Um outro aspeto possível de ser estudado, extremamente importante no estudo das vogais nasais, é a organização temporal dos gestos dos diferentes articuladores. A existência de uma fase oral na produção da vogal nasal e a emergência de um segmento consonântico nasal final podem ser explicados através da coordenação dos vários gestos: o gesto oral, o gesto velar e gesto da glote (Oliveira e Marin, 2005; Oliveira e Teixeira, 2007b; Oliveira, 2009). Se o gesto de abertura do véu palatino se iniciar depois do início do gesto vocálico, criam-se as condições para a existência de uma fase oral. Se existir sobreposição do gesto de fecho do velo com o *target* oral seguinte (fecho dos lábios no caso de uma oclusiva como o [p]), é possível a emergência de uma consoante nasal (Oliveira e Teixeira, 2007b; Oliveira, 2009). No que diz respeito ao gesto da glote, não será possível obter essa informação, uma vez que a qualidade das imagens a nível da glote não nos permite verificar a sua abertura ou fecho e consequentemente a existência ou não de vozeamento.

Para extrair os dados de coordenação entre os gestos dos diferentes articuladores será necessário ter informação relativa à variação temporal destes. Neste caso terá que se obter informação, de cada ponto temporal, relativa ao movimento do véu palatino e dos lábios. Assim, recorrendo ao método descrito no final da secção 6.4.5.2 obtiveram-se as curvas de variação do movimento destes articuladores ao longo do tempo. Na figura 6.29 apresenta-se um exemplo destas curvas, para cada uma vogais nasais, resultantes de produções da falante CO.

No tocante a este assunto, o tratamento dos dados é ainda muito preliminar resumindo-se à apresentação de alguns parâmetros temporais. Estes foram extraídos diretamente, de forma manual, a partir das curvas de variação do movimento dos articuladores. Apesar de não ser a forma mais eficiente de o fazer¹¹, pareceu-nos importante averiguar se seria possível retirar, a partir da base de dados de RM (com a resolução temporal atingida), informação semelhante e complementar à obtida através de EMA.

Neste sentido, seguimos a metodologia descrita por Oliveira (2009) e calculámos o *Target-to-Target Lag* (TTLag) que corresponde ao período temporal entre o *target* de abertura do véu

¹¹O processo que usualmente é utilizado, para o cálculo destes parâmetros, advém do cálculo da primeira derivada das curvas, sendo mais fácil definir os *landmarks* relativos às transições (abertura/fecho) entre os gestos. O método é semelhante ao que já utilizámos, neste trabalho, para extração dos valores de duração do gesto do véu palatino. Os valores temporais podem ser extraídos de forma automática.

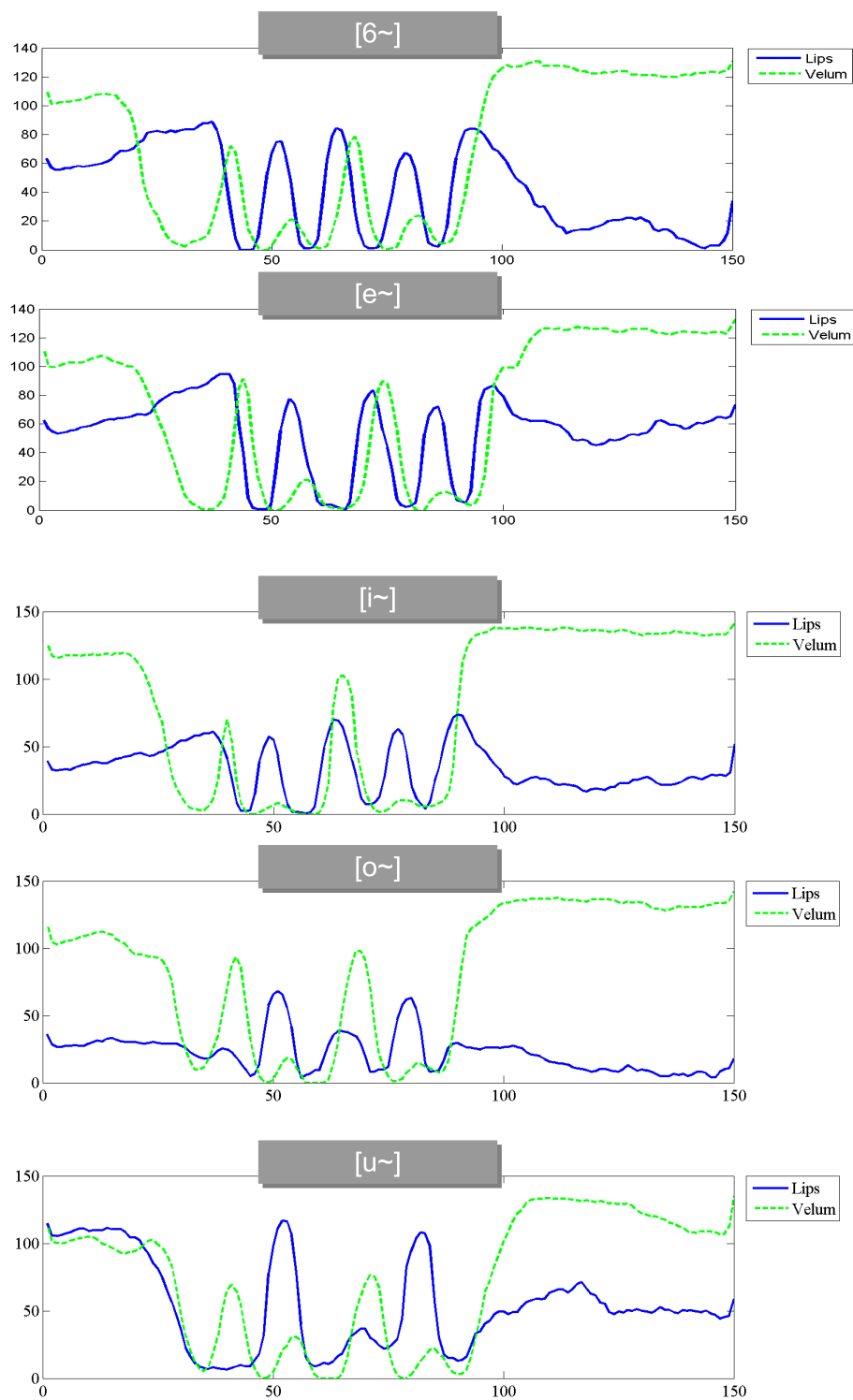


Figura 6.29: Curvas de variação do véu palatino e lábios, ao longo do tempo, para todas as vogais nasais, efetuadas pela falante CO.

palatino e o *target* do fecho oral - e o T2TL (*Target2-to-Target Lag*), que considera o *target* de fecho do véu palatino (T2) e o *target* de fecho oral.

Estes valores permitem obter informação relativa à coordenação entre a abertura do véu palatino e o fecho oral (TTLag) e entre o fecho do véu palatino e o fecho oral (T2TTLag). Para tal torna-se necessário identificar alguns pontos de referência (*landmarks*) sobre as curvas (e.g., máxima abertura do velo ou *target* de abertura). Na figura 6.30 apresenta-se um exemplo das marcações efetuadas e a forma com foram obtidos os parâmetros referidos. Sobre a curva do véu palatino (verde tracejado) foi marcado o (T) que representa o *target* de abertura do véu palatino e (T2) o *target* de fecho. Na curva dos lábios (azul) é identificado o *target* de fecho oral (T). No futuro, uma marcação mais precisa destes parâmetros temporais deverá passar pelo cálculo da derivada das curvas, uma vez que é mais fácil identificar as passagens por zero e os respetivos movimentos articulatorios.

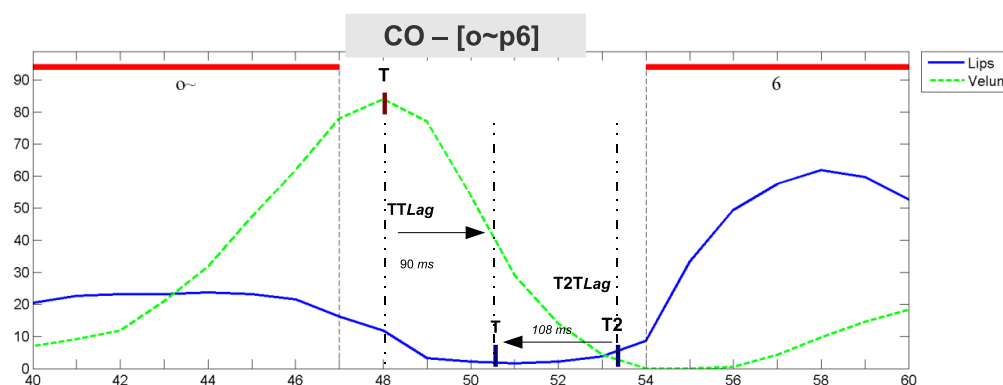


Figura 6.30: Medidas de sincronização entre o gesto do véu palatino e o gesto oral. Falante CO durante a produção de [o~] como produzido na palavra [o~p6].

Os resultados obtidos (para o TTLag e T2TTLag) resumem-se nas tabelas 6.6 e 6.7 respetivamente, e dizem respeito à falante CO, num total de doze produções com as vogais nasais em posição inicial e em posição medial.

Posição	Média (ms)	DP
Inicial	-77.5	24.9
Medial	-96.9	29.0

Tabela 6.6: Valores médios (ms) e desvio padrão da distância temporal entre o *target* de abertura do véu palatino e o *target* do gesto oral (TTLag), para as posições inicial e medial.

Posição	Média (ms)	DP
Inicial	117.5	33.0
Medial	119.4	24.5

Tabela 6.7: Valores médios (ms) e desvio padrão da distância temporal entre o *target* de fecho do véu palatino e o gesto de fecho oral (T2TTLag), para as posições inicial e medial.

Como se observa, os valores de TTLag são todos negativos o que significa que o véu palatino abre antes do fecho dos lábios. Estes resultados apontam para um tipo de coordenação

sequencial entre o gesto de abertura do véu palatino e o gesto oral, nas duas posições consideradas (inicial e medial).

As figuras 6.31 e 6.32 mostram exemplos deste tipo de coordenação para uma produção com a vogal em posição inicial como em [i~p6] e em posição medial como em [po~p6]. Os retângulos identificam o início e o final de cada um dos gestos (abertura e fecho) para cada um dos articuladores.

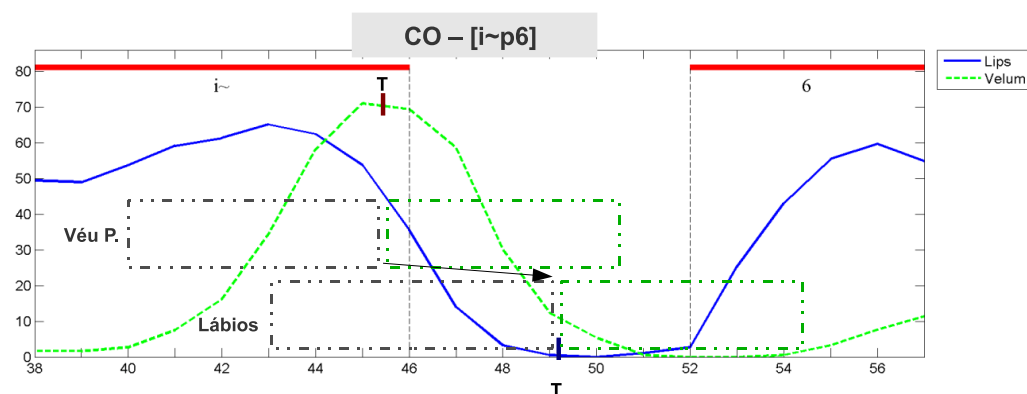


Figura 6.31: Exemplo de coordenação sequencial entre o gesto de abertura do véu palatino e o gesto de fecho labial, durante a produção de [i~p6].

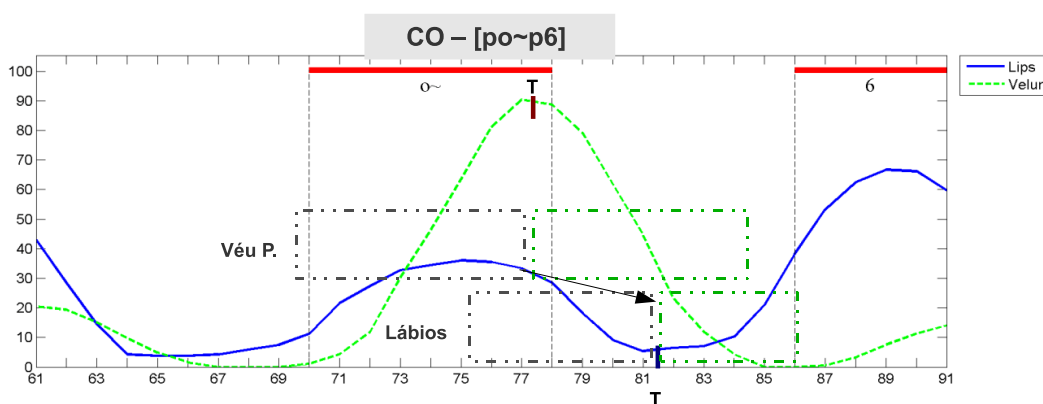


Figura 6.32: Exemplo de coordenação sequencial entre o gesto de abertura do véu palatino e gesto de fecho labial, durante a produção de [po~p6].

Dos aspetos de coordenação entre gestos e da maior ou menor sobreposição destes é que pode resultar a produção da vogal nasal em três fases: 1) fase oral em que o véu palatino está numa posição fechada ou quase fechada e os lábios abertos; 2) fase nasal propriamente dita em que o véu palatino está aberto e os lábios abertos e 3) um segmento nasal de natureza consonântica uma vez que o véu palatino está ainda bastante aberto e os lábios já atingiram o seu *target* (fechados). Estas três fases denotam o curso dinâmico da nasalidade no PE e foram observadas quando as vogais nasais foram produzidas em posição inicial e medial. Na figura 6.33 pretendemos ilustrar as diferentes fases da vogal nasal recorrendo a imagens de RM e às curvas de variação do movimento dos articuladores obtidas a partir destas. O exemplo diz respeito à vogal nasal [6~] produzida em

posição inicial pela falante CO.

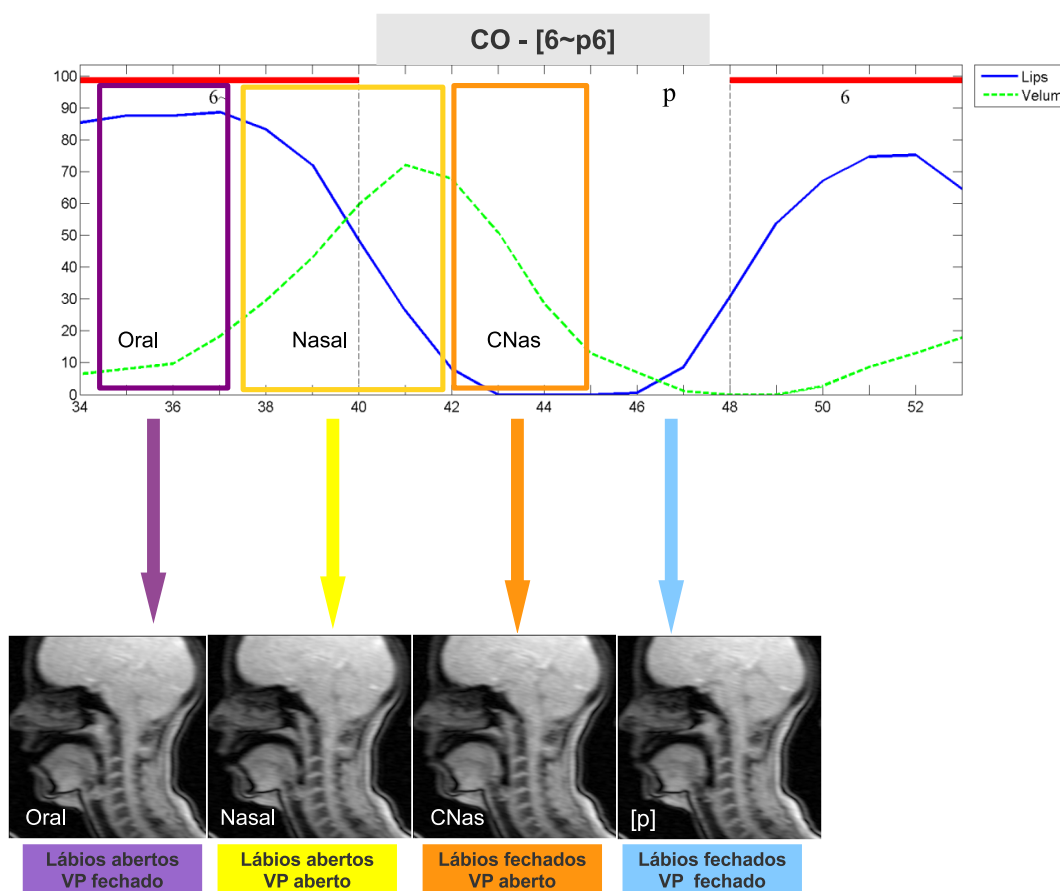


Figura 6.33: Identificação das três fases na produção das vogais nasais. Falante CO durante a produção de [6~] como produzido na palavra [6~p6].

Em posição final (cf. figura 6.34), o véu palatino e os lábios regressam à posição de repouso não sendo possível identificar o segmento consonântico nasal.

6.6 Discussão

Na presente secção, começaremos por efetuar uma análise crítica global relativamente ao que foi alcançado com esta parte do nosso trabalho, tendo em consideração os objetivos definidos. Depois focalizamo-nos nos resultados de produção obtidos fazendo o confronto com a literatura relacionada e fazendo a ponte com as questões enumeradas na secção 6.3.

6.6.1 Aspetos técnicos

No contexto deste trabalho, foi obtida uma base de dados de RM em tempo real sincronizada com o sinal acústico, para três falantes do PE, com um *corpus* ajustado, que no seu conjunto permitiu contribuir com mais informação para responder a várias questões relativamente à produção das

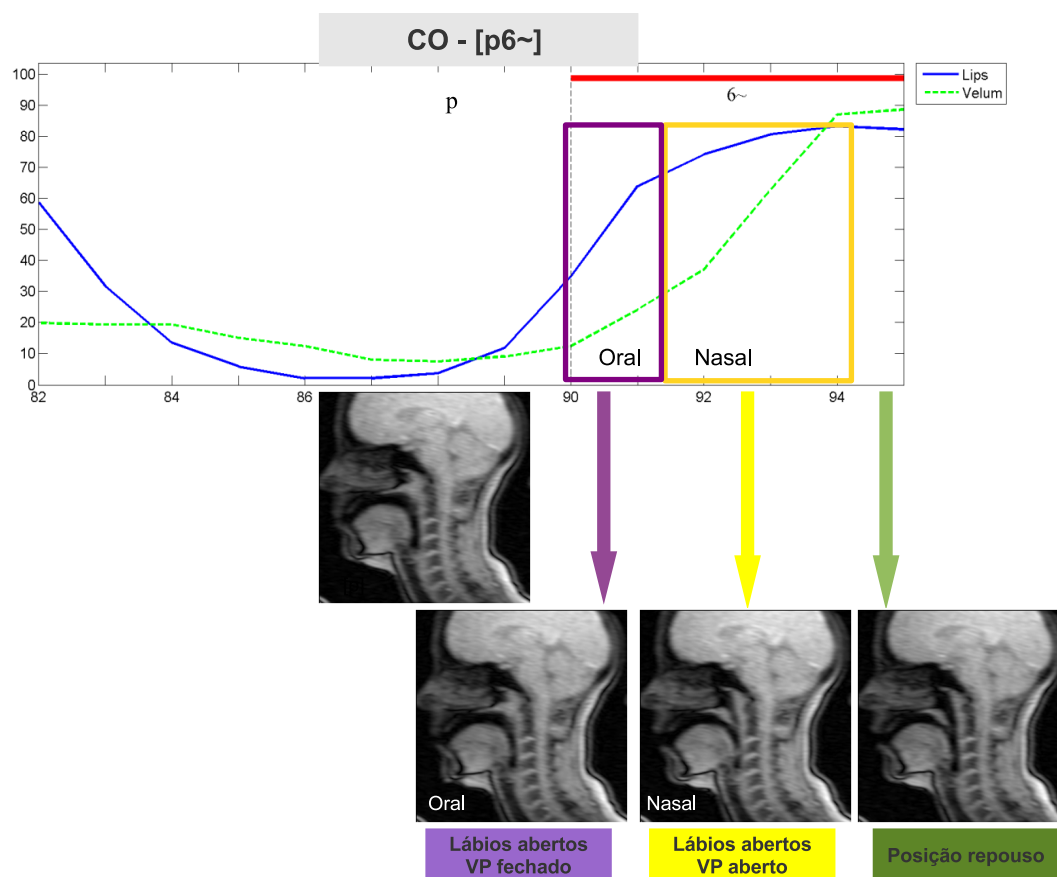


Figura 6.34: Identificação das diferentes fases na produção das vogais nasais. Falante CO durante a produção de [6 ~] como produzido na palavra [p6 ~].

vogais nasais do PE. Primeiramente, considerando os aspetos mais técnicos, é nossa convicção que os objetivos a esse nível foram plenamente atingidos:

No que diz respeito à resolução temporal que se conseguiu neste estudo em tempo real (72 *ms*), verificaram-se avanços muito significativos em relação ao que tinha sido a nossa experiência piloto com aquisição de imagens em tempo real em que foi obtida uma resolução temporal de 200 *ms* (Martins, 2007). A opção por uma outra sequência de pulsos, a aquisição de imagem num campo com maior intensidade, com gradientes de elevado desempenho, o recurso a imagem paralela viabilizada pela utilização de antenas *phased-array* contribuíram, indubitavelmente, para a melhoria da resolução temporal e da qualidade das imagens. As imagens obtidas não são degradadas por artefactos, permitindo uma boa visualização dos vários articuladores, particularmente, do véu palatino e da região velo-faríngea.

A qualidade das imagens obtidas possibilitou a utilização de rotinas de processamento bastante automatizadas, sendo necessários apenas ajustes manuais, de pormenor, o que se revelou extremamente importante para uma rápida obtenção dos resultados e para o processamento da informação desta base de dados relativamente extensa. Através das várias abordagens efetuadas, foi viável obter informação qualitativa e quantitativa relevante para um melhor conhecimento das características dinâmicas das vogais nasais.

A taxa de amostragem obtida revelou-se adequada para captar os movimentos do véu palatino e também dos outros articuladores (e.g., lábios e língua), ao longo do tempo, permitindo uma melhor caracterização dos gesto velar e oral, dando cumprimento a um dos nossos objetivos para este estudo. No que diz respeito aos aspetos de coordenação, apesar dos resultados a este nível ainda não terem sido completamente explorados, demonstrámos ser possível obter essa informação a partir das imagens de RM. Esta constatação vem corroborar as afirmações de Byrd *et al.* (2009) e de Proctor *et al.* (2013), no que diz respeito às potencialidades da RM neste domínio. Para além disso, foi possível, utilizando abordagens de análise similares, compatibilizar os dados de RM com os dados EMA previamente adquiridos, o que no seu conjunto resulta em mais informação relativa à articulação das vogais nasais do PE.

A realização dos cortes oblíquos na zona do véu palatino, explorando a capacidade multiplanar da RM, permitiu obter informação nova que vai para além da que é possível obter com as imagens sagitais ou através de dados EMA. Referimo-nos, em particular, à possibilidade de obter, de forma direta, as áreas nasais e orais na zona do acoplamento, ao longo do tempo, acompanhando a trajetória de deslocação do véu palatino. Acreditamos que esta abordagem permitirá vir a estabelecer comparações com dados obtidos através de outras técnicas como a fibroscopia, *velotrace* ou RM estática 3D, revisitando alguns temas com grande interesse para o estudo da nasalidade, como por exemplo as relações entre altura e abertura velar, distância médio-sagital e áreas na zona de acoplamento (veja-se, por exemplo, Lovatto *et al.* 2007, Demolin *et al.*, 2003 ou Serrurier e Badin, 2008).

Para além disso, os cortes coronais, permitem obter informação única relativa à parede posterior e laterais da faringe que, em conjunto com o véu palatino, delimitam o PVF. O acesso a esta zona (difícil para a maioria das técnicas), de forma não invasiva, pode ter um grande impacto,

por exemplo, no estudo da fala em indivíduos com alterações (e.g., hipernasalidade) e de outras patologias relacionadas com o mecanismo velo-faríngeo (Amelot *et al.*, 2003). De acordo com Kane *et al.* (2002) é relevante, nestes indivíduos, obter informação não só das áreas, mas relativa à própria configuração assumida pelo PVF, devido à ação das paredes laterais e da parede posterior da faringe.

6.6.2 Articulação das vogais nasais

Os resultados obtidos, que nos permitiram caracterizar os movimentos dos diversos articuladores envolvidos na produção das vogais nasais estão, em geral, de acordo com os resultados de trabalhos anteriores com outras técnicas (e.g., EMA). Sustentam também as propostas de Teixeira (2000), obtidas a partir de experiências de síntese de fala (SAPWindows), que apontam para: 1) maior naturalidade das vogais nasais sintetizadas quando se faz variar a posição do véu palatino ao longo do tempo e 2) a existência de um movimento inicial de fecho do velo condicionando um início oral da vogal nasal. Verifica-se também que os ajustes nos articuladores orais (e.g., língua), durante a produção das vogais nasais são pequenos, o que está em linha com resultados prévios, obtidos com RM estática, para o PE.

6.6.2.1 Movimentos do véu palatino

Em relação aos aspetos de duração dos movimentos de véu palatino, os resultados alcançados com a RM em tempo real são, no geral, muito próximos dos reportados para o PE através de EMA (Oliveira e Teixeira, 2007b; Oliveira, 2009), particularmente os obtidos para a falante CM. Estão também dentro dos valores reportados na literatura (para o ciclo completo dos movimentos do véu palatino) por outros autores (Stevens, 1998; Amelot, 2004). A duração total do movimento foi de 319.5ms (valor médio) para esta falante. Recorde-se que Oliveira (2009) obteve para a duração total do ciclo de abertura /fecho do véu palatino um valor médio de 320 ms. Para a falante CO, o valor médio da duração total foi um pouco mais elevado (cerca de 400 ms). Este resultado pode ser justificado pelo facto desta falante ter produzido as sequências a um ritmo mais lento. Na realidade, apesar de se ter solicitado a produção a um ritmo normal, este aspeto poderá ter variado um pouco entre as duas sessões de aquisição, não tendo sido adequadamente controlado. Apesar disso, verifica-se que a informação proveniente das imagens de RM, com a resolução temporal conseguida, é sensível a diferenças na taxa de elocução, viabilizando a sua utilização, no futuro, na avaliação de outros aspetos prosódicos.

Para além da duração total, os valores obtidos para a duração de cada uma das fases (abertura, fase estável e fecho) e a constatação de que a duração do movimento de abertura é superior ao de fecho (com excepção da vogal [u ~] em que se verifica o oposto) consubstanciam os dados EMA reportados, previamente, para o PE (Oliveira e Teixeira, 2007b; Oliveira, 2009) e para o inglês americano (Solé, 1995). Ou seja, nestas línguas o véu palatino abre lentamente e fecha rápido. Já no que diz respeito à língua francesa, de acordo com Amelot (2004), os valores de abertura e fecho tendem a ser semelhantes ou o de fecho superior ao de abertura.

6.6.2.2 Articulação oral das vogais nasais

No que diz respeito às vogais nasais altas ([i~] e [u~]), para além do abaixamento do véu palatino, não há diferenças relevantes nos articuladores orais comparativamente ao observado com as correspondentes orais, em concordância com o observado previamente por Martins *et al.* (2007) e Oliveira (2009), recorrendo a aquisições efetuadas com sustentação das vogais.

Comparando os resultados obtidos com os dados articulatórios existentes para o PB (provenientes de RM e xeroradiografia) verifica-se que há alguns aspetos comuns, nomeadamente o facto do [i~] ser das vogais em que os ajustes articulatórios orais são menores (Gregio, 2006; Master *et al.*, 1991). Para estas vogais a comparação com o Francês não pode ser estabelecida uma vez que nesta língua não existem vogais nasais com altura equivalente à das vogais altas do PE.

É de salientar ainda que quando se compararam os contornos obtidos, ao longo da produção de cada uma das vogais nasais foram também as vogais [u~] e [i~] as que apresentaram menos variação nos articuladores orais ao longo do tempo. Poder-se-á especular que, nas vogais altas, as condições aerodinâmicas para a produção da nasalidade estejam já presentes através de um maior fechamento no trato oral. A este respeito Master *et al.* (1991) refere que para além do abaixamento do véu palatino a nasalidade das vogais está “*intimamente relacionada com a constrição do trato vocal na cavidade oral.*”

As vogais [e~] e [o~] são as que apresentam maiores diferenças em relação às congéneres orais, mas ainda assim, quando adquiridas de forma dinâmica e em contexto, o que é um cenário mais realista, estas diferenças são menores que as observadas nos estudos estáticos conduzidos, anteriormente, para o PE (Martins *et al.*, 2007; Oliveira, 2009).

A sobreposição dos contornos não mostrou, em qualquer das produções das três falantes, a existência de ajustes a nível da parede posterior da faringe. Quando existiu estreitamento da faringe na produção da vogal nasal, este é à custa do recuo da língua e não aconteceu de forma sistemática, nem com a magnitude que tem sido reportado para o Francês ou mesmo para o PB como apontado por Matta Machado (1993). Na verdade, a única vogal nasal em que existe um maior recuo da língua em direção à faringe, dando origem a um estreitamento nessa zona, é o [e~].

Os ajustes orais entre os pares oral/nasal para o PE parecem ser mais subtis do que os reportados para o Francês (Delvaux *et al.*, 2002a), para vogais com altura semelhante às do PE. No entanto, deveremos considerar que, devido à natureza qualitativa das comparações efetuadas através da sobreposição de contornos, poderá existir alguma subjetividade na interpretação dos resultados.

Os resultados obtidos para as vogais produzidas de forma isolada e a sua comparação com as produções contextualizadas revelaram-se interessantes. Como se observou há mais alterações, ao longo do tempo, na produção de algumas vogais quando produzidas de forma isolada. Este facto pode ser explicado por uma tendência para hiperarticulação das vogais, quando produzidas de forma isolada ou porque o conjunto língua/mandíbula ficará, claramente, menos constrangido (podendo variar) quando não existe contexto adjacente. Pensamos também que o que observámos

para as vogais produzidas de forma isolada se enquadra nas descrições efetuadas por Gregio (2006) para o PB. Na realidade, no estudo de RM conduzido por esta autora, as vogais nasais foram produzidas de forma isolada e a autora refere uma fase final na produção da vogal nasal com “*movimento da língua*” condicionando uma constrição no trato oral. Isto poderá explicar, em parte, a existência destes ajustes para as vogais mais baixas e não ser necessário nas vogais altas ([i~] e [u~]) em que o trato oral já está mais constrito devido a uma posição elevada da língua na cavidade oral.

6.6.2.3 Variação das áreas no tempo

No que diz respeito ao que se analisou relativamente à variação das áreas nasais e orais ao longo da produção da vogal nasal, da área máxima do PVF e da proporção nasal/oral, os resultados permitem-nos constatar que:

1. Para todas as falantes, no início da produção da vogal nasal, as áreas nasais são muito pequenas, o que pressupõe que o véu palatino se encontra, nessa fase, numa posição elevada. Em contexto intervocálico (como em **pampa**), o facto de o véu palatino se encontrar fechado no início da vogal não é muito surpreendente e poderia ser explicado por fenómenos de coarticulação (Amelot e Rossato, 2007; Montagu, 2004). No entanto, constatou-se que, com excepção de uma falante (SV), mesmo em início de palavra, se verificou a mesma tendência. Como se demonstrou através da observação das imagens sagitais e das áreas, na zona de acoplamento, o véu palatino está aberto (na posição de repouso) e, em vez de iniciar o movimento de abertura a partir dessa posição, começa por fechar. Mesmo quando o véu palatino começa a abrir, as áreas são muito pequenas (cerca de $25mm^2$) e não serão suficientes para a perceção da nasalidade nessa fase da vogal nasal.

Teixeira e Vaz (2001), num estudo com EMA, descrevem este padrão para todas as vogais nasais, independentemente da sua posição na palavra. Os resultados obtidos por Oliveira (2009) apontam no mesmo sentido. O facto de a nasalidade poder não estar presente ao longo de toda a vogal nasal vai ao encontro dos resultados de testes perceptivos efetuados com ouvintes do PE. Stevens *et al.* (1987), num estudo que incluiu ouvintes portugueses, franceses e ingleses, verificaram que, contrariamente aos ouvintes do francês que preferem uma nasalidade forte e de longa duração, os ouvintes portugueses, para além do murmúrio nasal, têm preferência por estímulos mais nasalizados mas não necessariamente ao longo de toda a vogal (Stevens *et al.* (1987), p.4). Como se pode constatar, estudos de diferente natureza indicam que, de forma consistente para o PE, o início de uma vogal nasal é oral. Esta fase oral da vogal nasal tem sido atestada também para o PB como demonstram os estudos de Gregio (2006) com RM e de Lovatto *et al.* (2007) com fibroscopia.

2. No que diz respeito à máxima abertura do PVF, obtida através da área nasal máxima para cada uma das vogais nasais, verificou-se que estes valores não variaram muito um função da vogal nasal. Considerando que a máxima abertura do PVF se correlaciona com a posição do véu palatino (altura do véu palatino), arriscamos afirmar que as vogais nasais do PE

serão produzidas com uma altura do véu palatino equivalente. Ou seja, não seria necessário assumir diferentes *targets* para o véu palatino na modelação das diferentes vogais nasais, o que vai ao encontro dos resultados de Oliveira (2009) obtidos com EMA, para o PE.

Comparando com o Francês, os estudos de Delvaux *et al.* (2002a) e de Rossato *et al.* (2006) apontam para a existência de diferentes posições do véu palatino nas vogais nasais, em função da sua altura. Contudo, Demolin *et al.* (2003) assumem, a partir das áreas nasais na zona do PVF, que haverá apenas uma única posição de abaixamento do véu palatino para as vogais nasais do francês. Todavia, como no francês não existem vogais nasais altas os autores sugerem que, em línguas com existência de vogais nasais altas, isso poderá não acontecer. No português, como vimos, há vogais nasais altas ([i~] e [u~]) e os nossos resultados não apontam para a existência de diferenças dignas de registo, na abertura do PVF, entre estas e as vogais mais baixas.

Considerando os valores da área de abertura máxima do PVF, estes estão de acordo com os que têm sido reportados na literatura por outros autores (Stevens, 1998; Demolin *et al.*, 2003; Serrurier e Badin, 2008)¹². Demolin *et al.* (2003) verificaram, que a área do PVP varia bastante entre os sujeitos e que isso pode ser explicado por diferenças anatómicas.¹³ Note-se que, no nosso estudo, apesar da magnitude das áreas do PVF ser diferente para as três falantes, não se encontraram diferenças tão marcadas, entre os falantes, uma vez que eram todos do género feminino e com estatura similar. O estudo para o Francês incluía falantes do género masculino e feminino, o que pode conduzir a valores de abertura do PVF bastante diferentes entre os falantes.

3. Apesar de não se poder estabelecer uma comparação direta do P_{NO} obtido com os resultados de outros autores, estes estão de acordo com o que é a visão dinâmica da nasalidade vocálica defendida por Reenen (1982)¹⁴ e por Feng e Castelli (1996)¹⁵.

Nos nossos resultados, observou-se que o P_{NO} aumenta gradualmente ao longo da produção de qualquer uma das vogais nasais do PE, nunca atingindo o valor de 1 ou 100%. Os valores baixos no início poderão configurar uma fase oral da vogal nasal. O valor máximo é atingido na parte mais final da vogal e esse valor é diferente para cada uma das vogais nasais dependendo, como será óbvio, da dimensão da abertura oral.

No caso da vogal [u~], esse valor é próximo de 1 (0.99) que se deverá ao facto de ser uma vogal muito alta e muito recuada. Para o [o~], os valores são também altos ainda que menores (0.60 para a falante CO e 0.96 para a falante CM), pois apesar de recuada é uma vogal mais

¹²Por exemplo, Stevens (1998) refere que a máxima abertura do PVF se situa à volta de 1 cm^2 podendo ser maior. Na produção dos sons que envolvam a abertura do PVP as áreas variam entre 0.2 a 0.8 cm^2 . Demolin *et al.* (2003) para as falantes do género feminino referem valores entre 0.31 a 1.42 cm^2 na produção das várias nasais do Francês. Para os falantes masculinos estes valores são um pouco maiores, particularmente para um dos falantes. Note-se, todavia que há diferenças metodológicas entre o nosso estudo e os estudos referidos.

¹³Esta variabilidade entre falantes é também referida por Engwall *et al.* (2006), particularmente entre falantes femininos e masculinos.

¹⁴Tendo em atenção o $N\%$ proposto por Reenen (1982) a uma vogal oral pura corresponderia um $N\% = 0$ e as consoantes nasais teriam um $N\% = 100$ uma vez que neste caso há uma constrição no trato oral (área oral seria nula). As vogais nasais segundo a sua descrição seriam: "a dynamic phenomenon produced with an increase in $N\%$ from (almost) zero to about 75%" (Reenen, 1982, p.118).

¹⁵Estes autores referem que a configuração naso-faríngea será o *target* para o qual as vogais nasais tenderão a convergir ainda que nunca o atinjam.

baixa que o [u~]. Também teremos que admitir que para o [u~] a área oral possa ser um pouco maior do que a que foi medida, uma vez que devido ao tamanho do pixel (elevado) o algoritmo de segmentação não consiga resolver convenientemente uma abertura tão pequena. Para as vogais mais anteriores e para as mais baixas, o P_{NO} é mais baixo. Não encontramos qualquer correlação entre a altura da vogal e o grau de abertura nasal como pressuposto por Reenen (1982), em que uma vogal mais alta necessitaria de menos acoplamento nasal e vogais mais baixas um maior acoplamento. Na realidade, o que verificámos é que, para qualquer uma das vogais, incluindo o [u~], a abertura nasal é semelhante e a proporção varia, essencialmente, à custa da área oral. Tal como referido em Sampson (1999), muitos outros estudos não conseguiram estabelecer esta relação entre altura da vogal e a abertura do PVF.

Não tendo sido considerado no *corpus* adquirido, parece-nos muito indicado, em futuras aquisições, incluir a realização de cortes oblíquos na zona de acoplamento, não só para as vogais nasais, como foi efetuado, mas também para as vogais orais e consoantes nasais. Desta forma será possível obter os valores da proporção nasal/oral para estes sons e compará-los com as vogais nasais. Suspeitamos que essa proporção, para as vogais orais será mais pequena que para as nasais mas não será necessariamente zero como por vezes é considerado para estes segmentos orais (Reenen, 1982).

6.6.2.4 Aspetos de coordenação temporal/produção da vogal nasal em três fases

A interpretação dos resultados, compilando toda a informação recolhida, permite-nos apontar para a existência de três fases na produção das vogais nasais do PE: fase oral, fase nasal propriamente dita e um segmento nasal final. A existência de uma fase oral na produção das vogais nasais do PE é suportada por vários aspetos:

- quando se observaram as imagens de RM no plano sagital verificou-se que, na fase inicial da vogal nasal, o véu palatino está fechado e só começa a abrir posteriormente;
- obtiveram-se valores pequenos de área nasal no início da vogal nasal;
- a proporção nasal/oral é baixa no início da vogal nasal.

A existência desta fase oral tem sido referida por vários autores, utilizando técnicas de estudo diferentes, quer para o PE (Teixeira, 2000; Oliveira, 2009), quer para o PB (Souza, 1994; Seara, 2000; Gregio, 2006; Lovatto *et al.*, 2007; Medeiros, 2011).

Depois desse início oral, o véu palatino abre progressivamente ao longo da vogal até atingir o seu valor máximo de abertura na fase final da vogal nasal, configurando uma fase nasal propriamente dita. Os dados já coligidos (TTLag e T2TLag) apontam para a existência de um segmento nasal uma vez que ocorre, no final da produção da vogal nasal, um período em que os lábios já estão fechados (oclusão) ou seja atingiram o seu *target* e o véu palatino ainda está aberto, configurando uma consoante nasal. Estes resultados corroboram e confirmam as experiências efetuadas por Teixeira (2000) com síntese articulatória, os resultados obtidos com EMA por Oliveira

(2009) e os dados acústicos obtidos por Domingos (2011) e por Ashby *et al.* (2011) que também apontam para a existência de um segmento consonântico final no PE.

Ainda assim, consideramos que os dados de coordenação obtidos (TTLag e T2TTLag) necessitam de ser aferidos através do cálculo das derivadas das curvas de variação do velo e dos lábios. Só desta forma será possível obter valores mais exatos relativos, por exemplo, à duração deste segmento consonântico. Falta ainda, no tocante às vogais nasais do PE, analisar a parte da base de dados que inclui estes sons produzidos em contexto de fricativa lábio-dental [f], o que permitirá verificar se também, neste caso, estão criadas as condições para a emergência da consoante nasal final.

Em suma, para além da descrição das metodologias seguidas nas aquisições de imagem com RM em tempo real e seu processamento, apresentaram-se vários exemplos de aplicação desta técnica, permitindo obter informação bastante diversificada com contributos importantes para o estudo da articulação das vogais nasais do PE. A RM em tempo real revelou-se uma ferramenta versátil e poderosa para a obtenção dessa informação consubstanciando a opinião que tem vindo a ser formada relativamente à utilização desta modalidade imagiológica no domínio da produção de fala.

Conclusões

As ciências começam e acabam com um crepúsculo, com um ponto de interrogação.

PAOLO MONTEGAZZA

Neste capítulo, que encerra a presente tese, apresenta-se uma síntese do trabalho desenvolvido e os principais resultados e contributos que decorreram do mesmo. Finalmente, apontam-se áreas de interesse para desenvolvimento futuro, que podem dar continuidade ao trabalho agora efetuado e, também, ao trabalho desenvolvido, na Universidade de Aveiro, no âmbito da produção de fala.

7.1 Resumo do trabalho e principais contributos

O trabalho agora apresentado constitui uma extensão e evolução da investigação iniciada no âmbito do Mestrado em Ciências da Fala e Audição, concluído em 2007 (Martins, 2007), e dos trabalhos desenvolvidos por Carbone (2008) e Oliveira (2009), todos eles decorrentes de diferentes tarefas do Projeto HERON.

O presente trabalho dividiu-se em quatro grandes tarefas que incluíram: 1) otimização de protocolos/procedimentos 2) aquisição de bases de dados de produção de fala com recurso a RM, 3) processamento da informação adquirida e 4) análise de dados, com vista a poder contribuir, por um lado, para um melhor conhecimento da articulação dos sons do PE e, por outro, visando criar as condições que permitam a evolução dos modelos de síntese de fala de base articulatória.

Pelas razões já explicitadas ao longo do trabalho, os sons identificados como de maior interesse foram as consoantes laterais e as vogais nasais. Pelas suas diferentes características estes sons levaram à adopção de técnicas de aquisição e de processamento também elas diferentes.

Nesta tese, perspetivamos caminhos e apresentamos soluções relativas à aquisição e processamento de imagens de RM, no âmbito da produção de fala. Os pontos que a seguir se apresentam, esquematizados também na figura 7.1, resumem os principais contributos deste trabalho,

repartidos pelas diferentes áreas em que o nosso trabalho se dividiu e espelhando as suas principais tarefas.

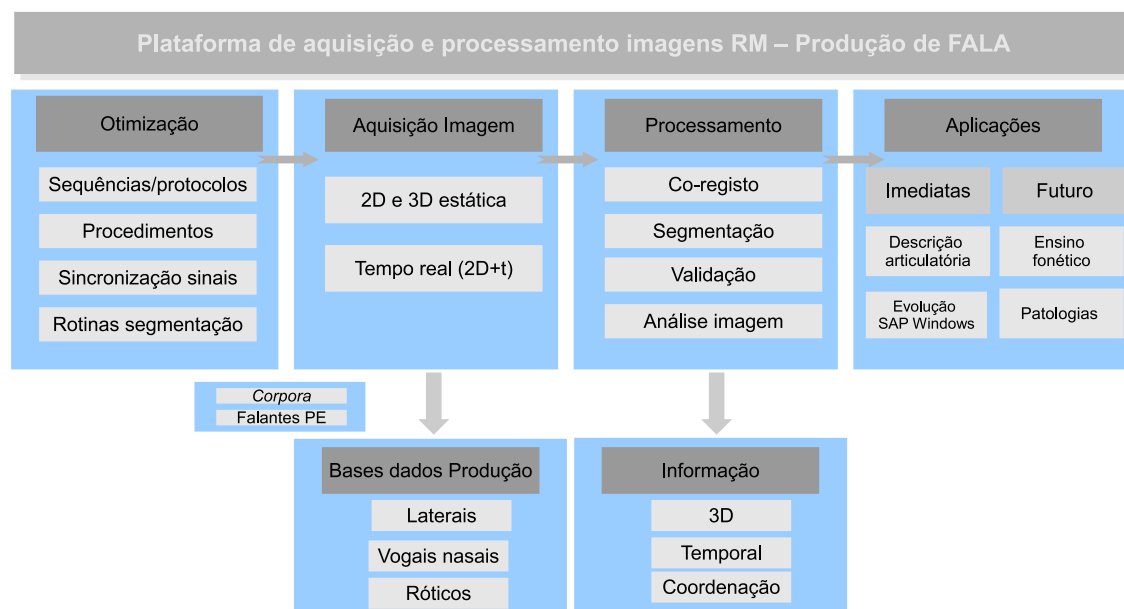


Figura 7.1: Resumo dos contributos da tese nas diferentes áreas do trabalho e o seu potencial de aplicação.

1. **Otimização** - Através de um processo de análise e experimentação, temos disponíveis, neste momento, protocolos de RM otimizados para responder aos vários e complexos requisitos inerentes ao estudo da produção de fala. Existem abordagens para aquisição de imagem 2D, 3D, tempo-real e possibilidade de sincronização de imagem com sinal acústico, tendo sido ultrapassadas algumas das limitações referidas em estudos anteriores (e.g., Teixeira, 2000; Martins, 2007 e Oliveira, 2009). Estes protocolos poderão permitir, de forma rápida e simplificada, a recolha de mais informação, sempre que sejam necessários dados de imagem para responder a questões para além daquelas que foram formuladas e exploradas no âmbito deste estudo.

Algumas das abordagens, por nós propostas (e.g., utilização da sequência VIBE 3D) já foram adotadas no seio de outras equipas de investigação para obter dados de produção para outras línguas como, por exemplo, o Árabe (Shosted *et al.*, 2012b).

2. **Novos dados produção** - No âmbito deste trabalho, foram criadas bases de dados de produção, para o PE, para as consoantes laterais (em 2D e 3D) e para as vogais nasais (em tempo real). Os *corpora* adquiridos contemplaram efeitos posicionais e contextuais para um número considerável de falantes, constituindo um importante avanço relativamente aos estudos prévios efetuados com RM, quer para o PE (e.g., Martins, 2007), quer para outras línguas (e.g., Narayanan *et al.*, 1997a). Para além destes sons, recolheram-se dados para as vogais centrais (2D e 3D) e para os sons róticos (2D, 3D e tempo-real). Relativamente a estes últimos pensamos que será um importante contributo, uma vez que também pouco se sabe relativamente à sua produção. Existindo dados já recolhidos e estratégias bem delineadas

para efetuar o seu processamento, rapidamente se obterá informação relevante para uma caracterização articulatória destes sons.

3. Contributos na área do Processamento de Imagem - Relativamente ao processamento de imagem, a autora contribuiu para a implementação de diversas rotinas de segmentação semi-automáticas (contributo mais relevante no processamento das imagens 3D), que permitem lidar com os vários tipos de imagens adquiridas (2D, 3D, tempo-real), constituindo-se como alternativas à segmentação manual e ultrapassando algumas das limitações reportadas, nesta área, por outros autores (e.g., Delvaux *et al.*, 2002a). O grau de automatização conseguida é variável, dependendo das estruturas, da qualidade das imagens, da informação pretendida e da maior ou menor complexidade geométrica do trato vocal na produção dos diferentes sons. Concretamente, neste domínio, foi efetuado:

- Co-registo das arcadas dentárias com os volumes obtidos em fonação, utilizando, exclusivamente, imagens de RM;
- Processamento e análise da informação adquirida para as consoantes laterais (e.g., língua 2D, em 3D, trato vocal) e parte dos dados obtidos para as vogais nasais (e.g., Martins *et al.*, 2011b ou Martins *et al.*, 2012b).
- Avaliação das segmentações, efetuadas para a língua, com recurso ao algoritmo baseado em regiões de competição, comparativamente a segmentações manuais feitas por um observador experiente, recorrendo a diferentes métricas (Martins *et al.*, 2011c). Esta avaliação permitiu verificar que esta abordagem poderá ter potencial de aplicação, também em contexto clínico, necessitando contudo de um processo de validação mais robusto. Por outro lado, as métricas a que recorreremos para efeitos da avaliação podem vir a ser utilizadas, por exemplo, para quantificar efeitos coarticulatórios em 3D, permitindo evoluir o trabalho já iniciado com informação 2D (veja-se como exemplo Martins *et al.*, 2011a);
- Aplicação, de forma inovadora, de ferramentas de computação gráfica para estudo dos sons da fala. A sobreposição de malhas em 3D e sua comparação pode ser utilizada para comparar sons, quantificar diferenças entre falantes, etc. (Martins *et al.*, 2011d, 2012a);

4. Contributos para a Fonética (Articulatória) do PE - Neste trabalho, apresenta-se uma descrição articulatória das consoantes **laterais**, ilustrada com imagens de RM e visualizações 3D que permitem (de forma intuitiva) perceber a forma e postura assumida pelos diversos articuladores na produção destas consoantes. Esta descrição é consubstanciada com outro tipo de informação qualitativa (e.g., contornos, sobreposição contornos) e dados quantitativos (e.g., extensão de contacto, áreas, funções de área, extensão e área das passagens laterais) fundamentais para a sua caracterização e para aplicação no processo de síntese destes sons. A informação agora obtida (Martins *et al.*, 2010; Oliveira *et al.*, 2011), em conjunto com dados EMA já existentes (Oliveira *et al.*, 2010), e com dados acústicos recolhidos recentemente (Marques, 2010; Monteiro, 2012), tendo por base um *corpus* de fala semelhante,

deu um contributo importante para a clarificação de algumas questões relativas à produção destes sons. De facto, os resultados apontam para um /l/ velarizado, independentemente da sua posição silábica; para a existência de alguns efeitos contextuais que são, todavia, idênticos para o /l/ produzido em posição inicial e final de sílaba. No que diz respeito à lateral palatal, os dados obtidos apontam para um ponto de articulação mais anterior do que tradicionalmente descrito para o PE e para uma área de contacto bastante extensa. Apesar de existirem diferenças entre as duas laterais, foram identificadas características comuns a ambas, estando estas, muito provavelmente, relacionadas com as condições aerodinâmicas que têm que ser satisfeitas para a sua produção e com o modo de articulação.

Relativamente às **vogais nasais**, foi possível obter novos dados de produção (e.g., variação de áreas ao longo do tempo) que, em conjunto com outra informação recolhida previamente, permitirá vir a melhorar a qualidade na síntese de sons nasais. O trabalho já iniciado por Oliveira (2009) com informação RM (estática), no que diz respeito ao estudo do gesto oral das vogais nasais, foi agora complementado com informação dinâmica (Oliveira *et al.*, 2012), com uma resolução que, para o véu palatino, se pode considerar em tempo real. Conseguiu-se também obter informação relativa aos movimentos de abertura/fecho do véu palatino, estando em linha com os dados EMA já existentes, agora para um maior número de falantes (Martins *et al.*, 2012b).

No que concerne aos aspetos de coordenação, pese embora os dados não tenham ainda sido totalmente analisados, demonstramos que esta informação pode ser obtida a partir das imagens de RM, à semelhança do efetuado com dados EMA, abrindo a possibilidade de congregar os dados obtidos com técnicas diferentes. No seu conjunto, os resultados obtidos acrescentam informação nova e sustentam/validam os resultados previamente alcançados. Os possíveis contributos, para a área da Fonética, não se esgotam com os resultados apresentados neste trabalho. Caberá agora aos foneticistas e linguistas interpretar e utilizar a informação agora disponibilizada.

Tendo em atenção o exposto nos pontos anteriores, podemos dizer que os objetivos deste trabalho foram globalmente conseguidos, existindo no presente mais informação sobre a produção dos sons que nos propusemos estudar (contributos fonéticos) e evoluções a nível técnico que no futuro propiciarão, de uma forma sistemática, a recolha de dados para mais falantes, para outros sons, de modo a abordar muitas das questões ainda em aberto relativas à produção dos sons da fala e grande potencial de aplicação noutras áreas.

No que diz respeito ao SAPWindows, pensamos que o trabalho por nós efetuado poderá dar um contributo importante para que essa evolução seja possível. Em particular, a modelação e síntese dos sons laterais virá a beneficiar com a informação proveniente dos dados articulatórios recolhidos no âmbito deste trabalho. Na realidade, o trabalho de evolução do sintetizador para um modelo 3D já está em curso (Silva, 2014). A informação recolhida relativa às vogais nasais também contribuirá para uma melhoria da qualidade na síntese destes sons.

É importante salientar que este trabalho só foi possível devido ao seu enquadramento no Projecto HERON II, uma vez que forneceu o suporte financeiro e logístico necessário para as

aquisições RM, para a aquisição de material (e.g., o microfone ótico) e para a divulgação dos resultados obtidos. O Projeto HERON II, como já foi referido anteriormente, é mais abrangente que o presente trabalho e visa a criação de um sistema completo de conversão de texto para fala, de base articulatória, para o Português. No que diz respeito a este projeto, uma das tarefas nele prevista - a recolha de dados RM - fica também cumprida e concluída.

A equipa multidisciplinar constituída no âmbito do anterior Projeto HERON está agora mais consolidada, permitindo a cooperação com outras equipas que trabalham em áreas afins. A oportunidade de trabalhar no seio de uma equipa alargada com profissionais de áreas distintas, revelou-se extremamente enriquecedora, permitindo à autora alargar horizontes em áreas como a fonética e linguística, processamento de imagem, etc.

Como técnica de radiologia, a realização deste trabalho contribuiu, indubitavelmente, para desenvolver e aprofundar conhecimentos dentro da nossa própria área de formação o que se considera de grande importância. O estudo da fala, utilizando a RM para obtenção de informação, constituiu um desafio exigente e permanente em que se procurou sempre encontrar o melhor compromisso para solucionar os vários problemas e obstáculos que se nos foram colocando ao longo do trabalho.

7.2 Trabalho Futuro

Para finalizar, gostaríamos de identificar algumas áreas em que nos parece ser possível dar continuidade e melhorar o trabalho agora apresentado. De entre as várias possibilidades destacaríamos as seguintes:

1. Terminar a análise dos dados já adquiridos no âmbito deste trabalho e que não foram ainda totalmente explorados. Referimo-nos, por exemplo, às vogais centrais, particularmente o [a] e o [ɨ]. Neste momento, existe informação RM (2D e 3D), para sete falantes havendo oportunidade de retomar e clarificar algumas questões relacionadas com a produção destes sons. Foram também adquiridos alguns dos sons róticos, quer em 3D, quer em tempo real que deverão ser alvo de análise;
2. Um dos problemas com que nos debatemos ao longo do estudo, que limitou de alguma forma a análise (sem tratamento estatístico robusto) e interpretação dos dados, está relacionado com a dificuldade em comparar e conjugar a informação proveniente de vários falantes. Também, as comparações através de sobreposição de contornos são subjetivas e difíceis de estabelecer. Neste sentido, há necessidade de procurar formas de normalização dos dados e recorrer a processos mais sistemáticos de análise da informação.
3. Estender à lateral-alveolar o trabalho de modelação acústica já iniciado para a palatal-lateral. Será importante perceber melhor o efeito da dimensão das passagens laterais e da cavidade supra-lingual no *output* acústico. No que diz respeito à modelação estão também reunidas as condições para que se possam vir a criar modelos 3D de vários articuladores como a língua, trato vocal ou véu palatino (e.g., modelação por elementos finitos);

4. Analisada a configuração articulatória 3D das laterais, o caminho a seguir será, a partir dos dados já adquiridos em tempo real, acrescentar a dimensão temporal, analisando a forma como os diversos articuladores se coordenam para a sua produção. Isto será particularmente importante para a lateral alveolar.
5. Uma vez que estão disponíveis, para o PE, dados articulatórios obtidos com EMA e se iniciou a recolha de dados com US, será de grande importância começar a co-registar esta informação, colmatando os pontos fracos e limitações de cada uma das técnicas;
6. Este trabalho, enquadrado na área da produção de fala, visou a aquisição de dados em indivíduos sem qualquer perturbação, no entanto, pensamos que grande parte das técnicas agora propostas e otimizadas constituiriam uma mais-valia se aplicadas em áreas de investigação mais clínicas ou mesmo na prática clínica. O desenvolvimento de investigação conjunta, por exemplo, com terapeutas da fala seria de grande interesse e poderia ter impacto no diagnóstico e monitorização de algumas patologias. Assim, neste domínio, destacam-se as seguintes:
 - Utilização conjunta de RM em tempo real e RM funcional, para verificar que alterações articulatórias e de planeamento são efetuadas por indivíduos com IVF ou com fendas palatinas na produção de sons nasais, tão frequentes na língua portuguesa. Estudos similares, com consoantes oclusivas, foram já divulgados com resultados muito promissores (Sato-Wakabayashi *et al.*, 2008). No caso de crianças com este tipo de patologias, a utilização de RM será com toda a certeza mais vantajosa e deverá ser considerada uma alternativa a técnicas que utilizam radiação ionizante (Drissi *et al.*, 2011; Perry *et al.*, 2011). A quantidade de estudos recentes que se focalizam na análise detalhada do mecanismo velofaríngeo através de RM 3D e tempo real é considerável. Apenas como exemplos apontamos os trabalhos de Tian *et al.* (2010), Bae *et al.* (2011) e Perry *et al.* (2013);
 - Outras áreas em que nos parece diretamente aplicável o trabalho agora desenvolvido é no âmbito de estudos relacionados com a apneia do sono para determinação dos volumes da coluna de ar ao nível da faringe, determinação das dimensões do próprio véu do palato, monitorização de intervenções, entre outra informação (Chan *et al.*, 2010; Kim *et al.*, 2013); em patologias da cavidade oral associadas com as dimensões da língua (Ajaj *et al.*, 2005) ou em glossetomizados (Zhou *et al.*, 2013).
7. Uma outra aplicação, bastante diferente das anteriores, que teria também muito interesse passaria pela utilização das imagens de RM, que agora existem para todos os sons do PE, como componente de apoio pedagógico. As imagens e seu tratamento com o recurso a técnicas de visualização e de computação gráfica adequadas (ver figura 7.2) permitirão a elaboração de manuais ilustrados que apoiem o ensino da fonética.

Considerando o desenvolvimento que se tem verificado na área da RM, parece-nos que o caminho a seguir passará agora por um investimento na melhoria da qualidade das aquisições 3D

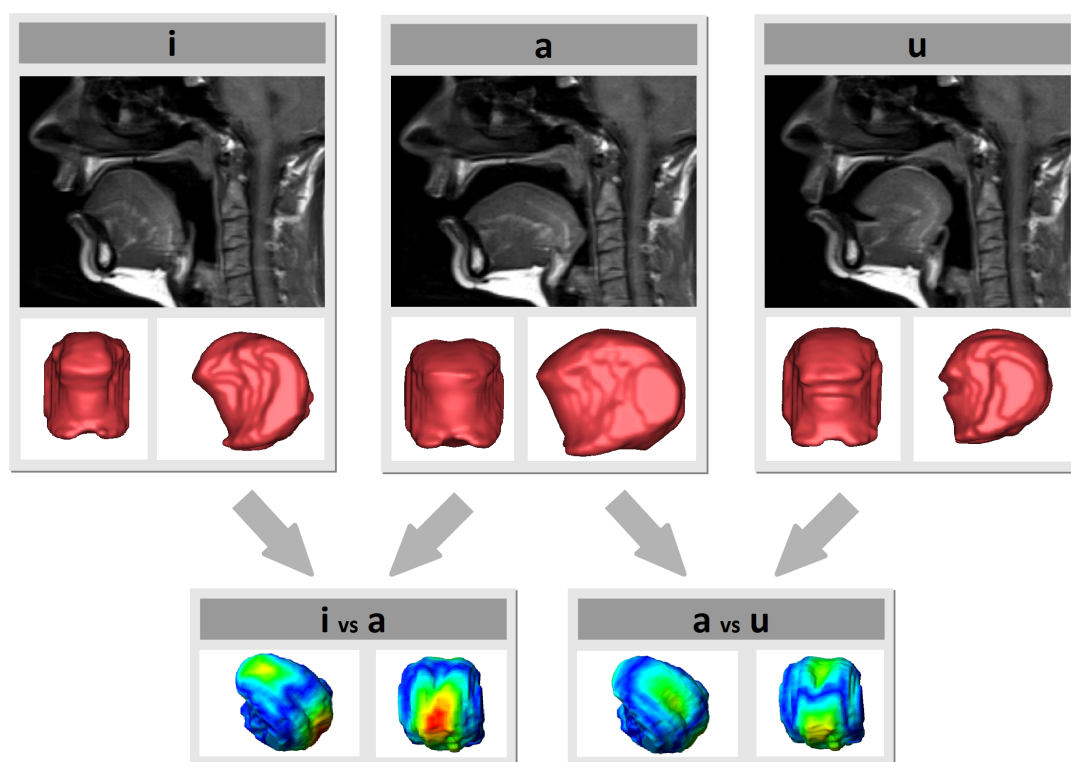


Figura 7.2: Exemplo de imagens que podem ser utilizadas para o ensino da fonética. Imagem publicada em Martins *et al.* (2012a)

em tempo real, congregando informação anatômica detalhada e informação temporal, não restrita a um plano. Esta possibilidade já se começa a delinear, tal como reportado, recentemente, em Zhu *et al.* (2013b) e Kim *et al.* (2013). Por outro lado, de forma concertada, seria importante começar a delinear estratégias de análise (sistemática) das muitas bases de dados (de RM e de outras técnicas) existentes para diferentes línguas.

7.3 Epílogo

*And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.*

THOMAS ELIOT

Bibliografia

- R. ADAMS E L. BISCHOF (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647.
- W. AJAJ, M. GOYEN, B. HERRMANN, S. MASSING, S. GOEHDE, T. LAUENSTEIN, E S. G. RUEHM (2005). Measuring tongue volumes and visualizing the chewing and swallowing process using real-time TrueFISP imaging - initial clinical experience in healthy volunteers and patients with acromegaly. *European Radiology*, 15(5):913–918.
- A. ALMEIDA (1976). The Portuguese nasal vowels: Phonetics and phonemics. *Readings in Portuguese Linguistics Amsterdam: North-Holland Publishing*, páginas 348–96.
- A. ALWAN, S. NARAYANAN, E K. HAKER (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *Journal of the Acoustical Society of America (JASA)*, 101(2):1078–1089.
- A. AMELOT (2004). *Etude aérodynamique, fibroscopique, acoustique et perceptive des voyelles nasales du français*. Tese de Doutorado, Université Paris III - Sorbonne Nouvelle, Paris.
- A. AMELOT, L. CREVIER-BUCHMAN, E S. MAEDA (2003). Observations of velopharyngeal closure mechanism in horizontal and lateral direction from fiberscopic data. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)- Barcelona, Spain*, páginas 3021–3024.
- A. AMELOT E S. ROSSATO (2007). Velar Movements for two French Speakers. In *Proceedings of the 16th International Congress of Phonetic Sciences(ICPhS)- Saarbrücken, Germany*, páginas 489–492.
- A. ANAGNOSTARA, S. STOECKLI, O. WEBER, E S. KOLLIAS (2001). Evaluation of the anatomical and functional properties of deglutition with various kinetic high-speed MRI sequences. *Journal of Magnetic Resonance Imaging*, 14(2):194–199.
- A. ANDRADE (1997). Variação Fonética do /l/ em Ataque Silábico em Português Europeu. In *Actas do XIII Encontro da Associação Portuguesa de Linguística*, páginas 55–76.

- A. ANDRADE (1999). On /l/ velarization in European Portuguese. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS) - San Francisco, USA*, páginas 543–546.
- A. ANDRADE E M. C. VIANA (1996). *Fonética*, páginas 115–167. Caminho, Lisboa.
- D. ARCHANGELI E J. BERRY (2010). Different dark and light /l/'s: English and Georgian. In *Proceedings of the Ultrafest V, New Haven, USA*.
- S. ASHBY, F. CUMMINS, S. BARBOSA, E N. CAMPANI (2011). Nasal heads or nasal tails? a pan lusophone survey of preconsonantal murmur. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)- Hong Kong, China*.
- M. AVILA-GARCÍA, J. CARTER, E R. DAMPER (2004). Extracting tongue shape dynamics from Magnetic Resonance Image sequences. *Transactions on Engineering, Computing and Technology*, 2:288–291.
- P. BADIN, G. BAILLY, M. RAYBAUDI, E C. SEGEBARTH (1998). A Three-Dimensional Linear Articulatory Model based on MRI Data. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP)- Sydney, Australia*, volume 2, páginas 417–420.
- P. BADIN, G. BAILLY, L. REVÉRET, M. BACIU, C. SEGEBARTH, E C. SAVARIAUX (2002). Three-dimensional Articulatory Modeling of Tongue, Lips and Face, based on MRI and Video Images. *Journal of Phonetics*, 30(3):533–553.
- P. BADIN, P. BOREL, G. BAILLY, L. REVÉRET, M. BACIUO, E C. SEGEBARTH (2000). Towards an Audiovisual Virtual Talking Head: 3D Articulatory Modeling of Tongue, Lips and Face Based on MRI and Video Images. In *Proceedings of the 5th International Seminar on Speech Production (ISSP) - Kloster Seeon, Bavaria, Germany*, páginas 461–464.
- P. BADIN, A. SERRURIER, *et al.* (2006). Three-dimensional linear modeling of tongue: Articulatory data and models. In *Proceedings of the 7th International Seminar on Speech Production, ISSP, Ubatuba, Brasil*, páginas 395–402.
- Y. BAE, D. P. KUEHN, C. A. CONWAY, E B. P. SUTTON (2011). Real-time magnetic resonance imaging of velopharyngeal activities with simultaneous speech recordings. *The Cleft Palate-Craniofacial Journal*, 48(6):695–707.
- T. BAER, J. C. GORE, S. BOYCE, E P. W. NYE (1987). Application of MRI to the analysis of speech production. *Mag. Res. Imag.*, 5(1):1–8.
- T. BAER, J. C. GORE, L. C. GRACCO, E P. W. NYE (1991). Analysis of vocal tract shape and dimensions using Magnetic Resonance Imaging: vowels. *Journal of the Acoustical Society of America (JASA)*, 90(2):799–828.
- P. BANGAYAN, A. ALWAN, E S. NARAYANAN (1996). From MRI and Acoustic Data to Articulatory Synthesis: A Case Study of the Lateral Approximants in American English. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP) - Philadelphia, USA*.

- I. BANKMAN (2000). *Handbook of Medical Imaging - Processing and Analysis*. Academic Press, New York.
- J. M. BARBOSA (1961). Les voyelles nasales portugaises: Interpretation phonologique. In A. SOVIJÄRVI E P. AALTO (editores), *International Congress of Phonetic Sciences (ICPhS)*, páginas 691–709. Mouton & Co.
- J. M. BARBOSA (1994). *Introdução ao Estudo da Fonologia e Morfologia do Português*. Livraria Almedina, Coimbra.
- J. BARKHAUSEN, M. GOYEN, F. WINTERFELD, T. LAUENSTEIN, D. ARWEILER-HARBECK, E J. F. DEBATIN (2002). Visualization of swallowing using real-time TrueFISP MR fluoroscopy. *European Radiology*, 12(1):129–133.
- J. BEHRENDTS E A. WISMÜLLER (2001). A segmentation and analysis method for MRI data of the human vocal tract. In *Proceedings of the FIPKM-37, Munich*, páginas 179–189.
- F. BELL-BERTI (1993). *Understanding Velic Motor Control: Studies of Segmental Context*, volume 5, páginas 63–85. Academic Press.
- F. BELL-BERTI, R. A. KRAKOW, D. ROSS, E S. HORIGUCHI (1993). *The rise and fall of the soft palate: The Velotrace*, páginas 21–22. Acoustical Society of America.
- A. P. BENGUEREL, H. HIROSE, M. SAWASHIMA, E T. USHIJIMA (1977). Velar coarticulation in french: Fiberscopic study. *Journal of Phonetics*, 5:149–158.
- M. BERNSTEIN, L. KING, E X. ZHOU (2004). *Handbook of MRI pulse sequences*. Elsevier Academic Press, Berlington.
- A. BILGIN, Y. KIM, H. G. LALGUDI, T. P. TROUARD, E M. I. ALTBACH (2008). Parallel Magnetic Resonance Imaging Using Compressed Sensing. In *Proceedings of SPIE - Applications of digital image processing XXXI, San Diego, USA*, volume 7073, páginas 70731G1–G10.
- P. BIRKHOLZ (2007). Articulatory synthesis of singing. In *Proceedings of Interspeech, Antwerp, Belgium*, páginas 4001–4004.
- P. BIRKHOLZ (2013). Modeling Consonant-Vowel Coarticulation for Articulatory Speech Synthesis. *PloS one*, 8(4):e60603.
- P. BIRKHOLZ, D. JACKÈL, E B. J. KRÖGER (2006). Construction and control of a three-dimensional vocal tract model. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)- Toulouse, França*, páginas 873–876.
- P. BIRKHOLZ E B. J. KRÖGER (2006). Vocal tract model adaptation using magnetic resonance imaging. In *Proceedings of the 7th International Seminar on Speech Production (ISSP) - Ubatuba, Brasil*, páginas 493–500.
- P. BIRKHOLZ, I. STEINER, E B. J. KRÖGER (2007). Control concepts for articulatory speech synthesis. In *Proceedings of the 6th ISCA Workshop on Speech Synthesis*, páginas 5–10.

- R. A. BLADON E A. AL-BAMERNI (1976). Coarticulation resistance of english /l/. *Journal of Phonetics*, 4:135–150.
- M. BLAIMER, F. BREUER, M. MUELLER, R. M. HEIDEMANN, M. A. GRISWOLD, E P. M. JAKOB (2004). SMASH, SENSE, PILS, GRAPPA: How to choose the optimal method. *Topics in Magnetic Resonance Imaging*, 15(4):223–236.
- P. BOERSMA E D. WEENINK (2009). Praat: doing phonetics by computer - [computer program]. version 5.1.
- D. BONE, M. PROCTOR, Y. KIM, E S. NARAYANAN (2011). Semi-automatic modeling of tongue surfaces using volumetric structural MRI. In *Proceedings of the 162nd Meeting of the Acoustical Society of America (ASA), San Diego, USA*.
- G. E. BOYLE, M. AHERN, J. COOKE, N. P. SHEEHY, E J. F. MEANEY (2006). An Interactive Taxonomy of MR Imaging Sequences. *Radiographics*, 26(6):e24.
- S. F. BRANDÃO (2007). Um estudo variacionista sobre a lateral palatal. *Letras de Hoje—Estudos e debates de assuntos de linguística, literatura e língua portuguesa*, 42(3):89–99.
- E. BRESCH (2011). *Visualizing and modeling vocal production dynamics*. Tese de Doutorado, University of Southern California - Faculty of USC Graduate School, USA.
- E. BRESCH, J. ADAMS, A. POUZET, S. LEE, D. BYRD, E S. NARAYANAN (2006a). Semi-automatic processing of real-time MR image sequences for speech Production Studies. In *Proceedings of the 7th International Seminar on Speech Production (ISSP)- Ubatuba, Brasil*.
- E. BRESCH, Y.-C. KIM, K. NAYAK, D. BYRD, E S. NARAYANAN (2008a). Seeing speech: Capturing vocal tract shaping using real-time Magnetic Resonance Imaging. *IEEE Signal processing magazine*, páginas 123–132.
- E. BRESCH E S. NARAYANAN (2009). Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images. *IEEE Transactions on Medical Imaging*, 28(3):323–338.
- E. BRESCH E S. S. NARAYANAN (2010). Real-time MRI investigation of resonance tuning in soprano singing. *J. Acoust. Soc. Am. Express Letters*, 128(5):335–341.
- E. BRESCH, J. NIELSEN, K. NAYAK, E S. NARAYANAN (2006b). Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans. *Journal of the Acoustical Society of America (JASA)*, 120(4):1791–1794.
- E. BRESCH, D. RIGGS, L. GOLDSTEIN, D. BYRD, S. LEE, E S. S. NARAYANAN (2008b). An analysis of vocal tract shaping in English sibilant fricatives using real-time magnetic resonance imaging. In *Proceedings of Interspeech, Brisbane, Australia*, páginas 2823–2826.
- F. A. BREUER, M. BLAIMER, M. F. MUELLER, N. SEIBERLICH, R. M. HEIDEMANN, M. A. GRISWOLD, E P. M. JAKOB (2006). Controlled aliasing in volumetric parallel imaging (2d

- caipirinha). *Magnetic Resonance in Medicine*, 55(3):549–556.
URL <http://dx.doi.org/10.1002/mrm.20787>
- T. BREYER, E. MATHIAS, S. ARNDT, B. RICHTER, O. SPECK, M. SCHUMACHER, E M. MARKL (2009). Dynamic Magnetic Resonance imaging of swallowing and laryngeal motion using parallel imaging at 3 T. *Magnetic Resonance Imaging*, 27:48–54.
- G. BRIX, H. KOLEM, E W. NITZ (2008). *Magnetic Resonance Tomography*, capítulo 2, páginas 3–167. Springer-Verlag.
- C. P. BROWMAN E L. GOLDSTEIN (1995). Gestural syllable position effects in American English. *Producing speech: Contemporary issues*, páginas 19–33.
- M. A. BROWN E R. C. SEMELKA (2010). *MRI: Basic Principles and Applications*. John Wiley and Sons, Inc, quarta edição.
- D. BYRD, S. TOBIN, E. BRESCH, E S. NARAYANAN (2009). Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics*, 37(1):97–110.
- L. C. CAGLIARI (1977a). Airflow in palatal laterals. *Work in Progress, Edinburgh University, Department of Linguistics*, 10:113–115.
- L. C. CAGLIARI (1977b). *An Experimental Study of Nasality with Particular Reference to Brazilian Portuguese*. Tese de Doutorado, University of Edinburgh, Scotland.
- H. CAMPOS (2009). *Duração dos segmentos vocálicos orais, nasais e nasalizados do português brasileiro*. Tese de Mestrado, Universidade Federal de Minas Gerais, Belo Horizonte.
- I. C. CARBONE (2008). *Segmentação do tracto vocal a partir de Estudos Imagiológicos de Ressonância Magnética*. Tese de Mestrado, Universidade de Aveiro, Aveiro.
- C. CARIGNAN (2011). Oral articulation of nasal vowels in French. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS), Hong Kong, China*, páginas 365–369.
- C. CARIGNAN, R. SHOSTED, C. SHIH, E P. RONG (2011). Compensatory articulation in American English nasalized vowels. *Journal of Phonetics*, 39(4):668–682.
- T. CARROLL, K. SAKAIE, P. WIELOPOLSKI, E R. EDELMAN (2006). *Advanced Imaging Techniques, Including Fast Imaging*, volume 1. Philadelphia: Elsevier.
- P. CARTER (2002). *Structured variation in British English liquids: The role of resonance*. Tese de doutoramento, Department of Linguistics, University of York, York.
- V. CASELLES, R. KIMMEL, E G. SAPIRO (1997). Geodesic Active Contours. *International Journal of Computer Vision*, 22(1):61–79.
- A. S. CHAN, K. SUTHERLAND, R. J. SCHWAB, B. ZENG, P. PETOCZ, R. W. LEE, M. A. DARENDELILER, E P. A. CISTULLI (2010). The effect of mandibular advancement on upper airway structure in obstructive sleep apnoea. *Thorax*, 65(8):726–732.

- G. B. CHAVHAN, P. S. BABYN, B. G. JANKHARIA, H.-L. M. CHENG, E M. M. SHROFF (2008). Steady-State MR Imaging Sequences: Physics, Classification, and Clinical Applications. *Radiographics*, 28(4):1147–1160.
- N. CHEN, J. SLIFKA, E K. STEVENS (2007). Vowel nasalization in American English: Acoustic Variability due to Phonetic Context. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS) - Saarbrücken, Germany*.
- W. CHEN E C. H. MEYER (2006). Fast automatic linear off-resonance correction method for spiral imaging. *Magnetic resonance in medicine*, 56(2):457–462.
- T. CHIBA E M. KAJIYAMA (1941). The vowel: Its nature and structure. *Tokyo-Kaiseikan Publishing Company, Ltd*.
- D. G. CHILDERS E C. DING (1991). Articulatory synthesis: nasal sounds and male female voices. *Journal of Phonetics*, (19):453–464.
- CIOMS (2002). *International ethical guidelines for biomedical research involving human subjects*. Council for International Organizations of Medical Sciences - World Health Organization.
- G. CLARKE (2007). Parallel Imaging: Techniques, Quality Control, and Applications (Preview of TG118 Report). *Medical Physics*, 34:2509.
- H. CLUMECK (1976). Patterns of soft palate movements in six languages. *Journal of Phonetics*, 4:337–351.
- I. COHEN, L. D. COHEN, E N. AYACHE (1992). Using Deformable Surfaces to Segment 3-D Images and Infer Differential Structures. *Computer Vision, Graphics and Image Processing: Image Understanding*, 56(2):242–263.
- L. D. COHEN E I. COHEN (1993). Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147.
- L. COLANTONI (2004). Reinterpreting the CV Transition: Emergence of the glide as an allophone of the palatal lateral. In C. CLEMENTS (editor), *Contemporary Approaches to Romance Linguistics*, páginas 83–102. John Benjamins.
- T. CRISTÓFARO-SILVA E M. OLIVEIRA (2001). Lateral Vocalization in Brazilian Portuguese. In *3rd UKLVC-University of York*, páginas 29–44.
- M. CRUZ-FERREIRA (1999). *Intonation in European Portuguese*. Cambridge University Press, Cambridge.
- C. CUNHA E L. F. L. CINTRA (1997). *Nova Gramática do Português Contemporâneo*. Edições João Sá da Costa, Lisboa, 13ª edição.
- B. M. DALE, H. SCHULTZE-HAAKH, E A. PRIATNA (2006). VIBE: Fast Gradient Echo imaging for every part of the body. *MagnetonFLASH*, 3:56–59.

- M. DEL FRESNO, M. VÉNERE, E A. CLAUSSE (2009). A combined region growing and deformable model method for extraction of closed surfaces in 3D CT and MRI scans. *Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society*, 33(5):369–376.
- P. DELATTRE (1968). Divergences entre nasalités vocalique et consonantique en Français. *Word*, 24:64–73.
- V. DELVAUX (2003). *Contrôle et connaissance phonétique: Les voyelles nasales du français*. Tese de Doutorado, Université Libre de Bruxelles, Bruxelles.
- V. DELVAUX, T. METENS, E A. SOQUET (2002a). French nasal vowels: acoustic and articulatory properties. In *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP), Denver, USA*, volume 1, páginas 53–56.
- V. DELVAUX, T. METENS, E A. SOQUET (2002b). Propriétés acoustiques et articulatoires des voyelles nasales du français. In *Proceedings of the XXIV èmes Journées d' Étude sur la Parole, Nancy*, páginas 348–352.
- D. DEMOLIN, V. DELVAUX, T. METENS, E A. SOQUET (2003). Determination of velum opening for french nasal vowels by magnetic resonance imaging. *Journal of Voice*, 17(4):454–467.
- D. DEMOLIN, M. GEORGE, V. LECUIT, T. METENS, A. SOQUET, E H. RAEYMAEKERS (1997). Coarticulation and articulatory compensations studied by dynamic MRI. In *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech)*, páginas 31–34.
- D. DEMOLIN, V. LECUIT, T. METENS, B. NAZARIAN, E A. SOQUET (1998). Magnetic resonance Measurements of the Velum Port Opening. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)- Sydney, Austrália*.
- D. DEMOLIN, T. METENS, E A. SOQUET (1996). Three-dimensional measurement of the vocal tract by MRI. In *4th International Conference on Spoken Language Processing (ICSLP)*, volume 1, página 272.
- D. DEMOLIN, T. METENS, E A. SOQUET (2000). Real Time MRI and Articulatory Coordinations in Vowels. In *Proceedings of 5th International Seminar on Speech Production (ISSP), Kloster Seeon, Bavaria, Germany*, páginas 86–93.
- A. DESHMANE, V. GULANI, M. A. GRISWOLD, E N. SEIBERLICH (2012). Parallel MR imaging. *Journal of Magnetic Resonance Imaging*, 36(1):55–72.
- O. DIETRICH (2007a). *Parallel imaging in clinical MR applications*, capítulo Single-Shot Pulse Sequences, páginas 119–126. Springer Berlin/Heidelberg.
- O. DIETRICH (2007b). *Parallel imaging in clinical MR applications*, capítulo General Advantages of Parallel Imaging, páginas 173–176. Springer Berlin/Heidelberg.

- O. DIETRICH (2007c). *Parallel imaging in clinical MR applications*, capítulo Limitations of Parallel Imaging, páginas 177–179. Springer Berlin/Heidelberg.
- M. DOMINGOS (2011). *Nasalidade Vocálica em Português- Pistas para identificação forense de falantes*. Tese de Mestrado, Faculdade de Letras, Universidade de Lisboa, Lisboa.
- C. DRISSI, M. MITROFANOFF, C. TALANDIER, C. AND FALIP, V. LE COULS, E C. ADAMSBAUM (2011). Feasibility of dynamic MRI for evaluating velopharyngeal insufficiency in children. *European Radiology*, 21(7):1462–1469.
- J. P. EARLS, V. B. HO, T. K. FOO, E. CASTILLO, E S. D. FLAMM (2002). Cardiac MRI: recent progress and continued challenges. *Journal of magnetic resonance imaging*, 16(2):111–127.
- M. ECHTERNACH, J. SUNDBERG, S. ARNDT, T. BREYER, M. MARKL, M. SCHUMACHER, E B. RICHTER (2008). Vocal tract and register changes analysed by real-time MRI in male professional singers—a pilot study. *Logopedics Phoniatrics Vocology*, 33(2):67–73.
- M. ECHTERNACH, J. SUNDBERG, S. ARNDT, M. MARKL, M. SCHUMACHER, E B. RICHTER (2010). Vocal tract in female registers—a dynamic real-time MRI study. *Journal of Voice*, 24(2):133–139.
- R. EDELMAN, E. DUNKLE, I. KOKTZOGLOU, A. GRIFFIN, E. RUSSELL, W. ANKENBRAND, A. RAGIN, E A. CARRILLO (2009). Rapid whole-brain magnetic resonance imaging with isotropic resolution at 3 Tesla. *Investigative Radiology*, 44(1):54–59.
- A. ELSTER E J. BURDETTE (2001). *Questions & Answers in Magnetic Resonance Imaging*. Mosby Inc, St. Louis, segunda edição.
- A. EMILIANO (2009). *Fonética do Português Europeu: Descrição e Transcrição*. Guimarães Editores, Lisboa.
- O. ENGWALL (2000a). A 3D tongue model based on MRI data. In *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP)*, volume III, páginas 901–904.
- O. ENGWALL (2000b). Are static MRI measurements representative of dynamic speech? Results from a comparative study using MRI, EPG and EMA. In *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP)*, volume I, páginas 17–20.
- O. ENGWALL (2002a). Evaluation of a System For Concatenative Articulatory Visual Speech Synthesis. In J. H. L. H. PELLOM E BRYAN (editores), *Proceedings of Interspeech, Denver, USA*, páginas 665–668.
- O. ENGWALL (2002b). *Tongue Talking: studies in intraoral speech synthesis*. Tese de Doutorado, KTH - Royal Institute of Technology.
- O. ENGWALL (2003). A revisit to the Application of MRI to the Analysis of Speech Production - Testing our assumptions. In *Proceedings of the 6th International Seminar on Speech Production, Sydney, Austrália*, páginas 43–48.

- O. ENGWALL (2004). From real-time MRI to 3D tongue movements. In *Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP)*, Jeju, Coreia.
- O. ENGWALL E P. BADIN (1999). Collecting and analysing two and three dimensional MRI data for Swedish. *Speech Transmission Laboratory: Quarterly Progress and Status Report (STL-QPSR)*, páginas 11–38.
- O. ENGWALL E P. BADIN (2000). An MRI study of Swedish fricatives: coarticulatory effects. In *Proceedings of the 5th Seminar on Speech Production, Kloster Seeon, Bavaria, Germany*, páginas 297–300.
- O. ENGWALL, V. DELVAUX, E T. METENS (2006). Interspeaker variation in the articulation of french nasal vowels. In *Proceedings of the 7th International Seminar on Speech Production, Ubatuba, Brasil*, páginas 3–10.
- C. ERICSDOTTER (2005). *Articulatory-Acoustic Relationships in Swedish Vowel Sounds*. Tese de doutoramento, Stockholm University.
- C. Y. ESPY-WILSON (1992). Acoustic Measures for Linguistic Features distinguishing the semi-vowels /wɹl/ in American English. *Journal Acoustical Society of America*, 92:736–757.
- EURATOM (1997). Diretiva 43/97 de 30 de junho. Relatório técnico, European Atomic Energy Community–EURATOM.
- A. FALCÃO (1997). *Pardigmas de segmentação de imagens guiadas pelo usuário: Live-wire, Live-lane e 3D Live-wire*. Tese de doutoramento, Universidade Estadual de Campinas - Faculdade de Engenharia Elétrica e Computação, Brasil.
- A. X. FALCÃO, J. K. UDUPA, E F. K. MIYAZAWA (2000). An Ultra-Fast User-Steered Image Segmentation Paradigm: Live Wire on the Fly. *IEEE Transactions on Medical Imaging*, 19(1):55–62.
- G. FANT (1960). *Acoustic Theory of Speech Production*. Mouton and Co., Gravenhage, The Netherlands, segunda edição.
- G. FANT (2001). On the speech code. *Quarterly Progress and Status Report TMH-QPSR*, 42(1).
- FDA (2003). Criteria for significant risk investigations of magnetic resonance diagnostic devices. Relatório técnico, Food & Drug Administration - Center for Devices and Radiological Health, USA.
- S. FELS, I. STAVENESS, A. HANNAM, J. LLOYD, P. ANDERSON, C. BATTY, H. CHEN, E C. COMBE (2009). Advanced Tools for Biomechanical Modeling of the Oral, Pharyngeal, and Laryngeal Complex. In *Proceedings of the International Symposium on Biomechanics, Healthcare, and Information Science, Kanazawa, Japan*.
- S. FELS, F. VOGT, K. VAN DEN DOEL, J. LLOYD, E O. GUENTER (2005). Artisynth: Towards realizing an extensible, portable 3d articulatory speech synthesizer. In *Proceedings of the International Workshop on Auditory Visual Speech Processing*, páginas 119–124.

- M. FENCHEL, U. HELBER, O. P. SIMONETTI, N. I. STAUDER, U. KRAMER, C.-N. NGUYEN, J. P. FINN, C. D. CLAUSSEN, E S. MILLER (2004). Multislice first-pass myocardial perfusion imaging: Comparison of saturation recovery (SR)-TrueFISP-two-dimensional (2D) and SR-TurboFLASH-2D pulse sequences. *Journal of Magnetic Resonance Imaging*, 19(5):555–563.
- G. FENG E E. CASTELLI (1996). Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization. *Journal of the Acoustical Society of America*, 99(6):3694–3706.
- L. FENG, M. B. SRICAI, R. P. LIM, A. HARRISON, W. KING, G. ADLURU, E. V. DIBELLA, D. K. SODICKSON, R. OTAZO, E D. KIM (2012). Highly accelerated real-time cardiac cine MRI using k-t SPARSE-SENSE. *Magnetic Resonance in Medicine*.
- P. FERNANDES (2008). *Modelação e Análise da Fusão Intersomática Cervical*. Tese de Mestrado, Instituto Superior Técnico - Universidade Técnica de Lisboa, Lisboa.
- W. T. FITCH E J. GIEDD (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106:1511–1522.
- A. K. FOLDVIK, U. KRISTIANSEN, E J. KVAERNESS (1993). A time-evolving three-dimensional vocal tract model by means of magnetic resonance imaging (MRI). In *Proceedings of the Third European Conference on Speech Communication and Technology*.
- J. B. FREIRE (2011). *Variação da Lateral Palatal na Comunidade de Jacaraú (Paraíba)*. Tese de Mestrado, Universidade Federal da Paraíba (UFPB), João Pessoa, Brasil.
- H. FUJITA, X. YANG, E T. ZHENG (2006). Array Coil Types and Design Principles. In *Proceedings of 14th Annual Meeting of the International Society Of Magnetic Resonance in Medicine ISMRM, Seattle, Washington, USA*.
- M. GARCIA (2010). O Segmento Lateral /l/ em Rima Interna. Sonoridade e Nuclearização em Português Europeu. *Revista de Estudos de Língua da Universidade de Porto*, 5:53–69.
- G. GERIG, M. JOMIER, E M. CHAKOS (2001). Valmet: A new validation tool for assessing and improving 3d object segmentation. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2001*, páginas 516–523. Springer.
- B. GICK (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. In J. LOCAL, R. OGDEN, E R. TEMPLE (editores), *Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation*, páginas 222–236. Cambridge University Press, Cambridge.
- B. GICK, F. CAMPBELL, S. OH, E L. TAMBURRI-WATT (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, 34:49–72.
- B. GICK, A. M. KANG, E D. H. WHALEN (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics*, 30(3):357–371.

- S. B. GILES E K. L. MOLL (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica*, 31:206–227.
- G. E. GOLD, R. F. BUSSE, C. BEEHLER, E. HAN, A. C. S. BRAU, P. J. BEATTY, E C. F. BEAULIEU (2007). Isotropic MRI of the Knee with 3D Fast Spin-Echo Extended Echo-Train Acquisition (XETA): Initial Experience. *American Journal of Roentgenology*, 188(5):1287–1293.
URL <http://www.ajronline.org/content/188/5/1287.abstract>
- U. GOLDSTEIN (1980). *An articulatory model for the vocal tracts of growing children*. Tese de Doutoramento, Massachusetts Institute of Technology, USA.
- S. GÖRGÜLÜ, D. SAĞDIÇ, E. AKIN, Ş. KARAÇAY, E N. BULAKBASI (2011). Tongue movements in patients with skeletal Class III malocclusions evaluated with real-time balanced turbo field echo cine magnetic resonance imaging. *American Journal of Orthodontics and Dentofacial Orthopedics*, 139(5):405–414.
- T. GOTO, S. NISHIDA, Y. NAKAMURA, K. TOKUMORI, Y. NAKAMURA, E K. KOBAYASKI (2007). The accuracy of 3-dimensional magnetic resonance 3D vibe images of the mandible:an in vitro comparison of magnetic resonance imaging and computed tomography. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, 103:550–559.
- J. GÉRARD, P. PERRIER, Y. PAYAN, *et al.* (2006). 3D biomechanical tongue modeling to study speech production. *Speech Production: Models, Phonetic Processes, and Techniques*, páginas 85–102.
- A. R. GREENWOOD, C. C. GOODYEAR, E P. A. MARTIN (1992). Measurement of vocal tract shapes using magnetic resonance imaging. *IEEE Signal processing magazine*, 139(6):553–560.
- F. N. GREGIO (2006). *Configuração do trato vocal supraglótico na produção das vogais do português brasileiro: dados de imagens de Ressonância Magnética*. Tese de Mestrado, Pontifícia Universidade Católica de São Paulo, São Paulo.
- M. A. GRISWOLD, P. M. JAKOB, R. M. HEIDEMANN, M. NITTKA, V. JELLUS, J. WANG, B. KIEFER, E A. HAASE (2002). Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210.
URL <http://dx.doi.org/10.1002/mrm.10171>
- H. GUERREIRO E S. FROTA (2010). Os processos fonológicos na fala da criança de cinco anos: tipologia e frequência. *Lisboa, Instituto de Ciências da Saúde, UCP*.
- M. HAACKE, R. BROWN, M. THOMPSON, E R. VENKATESEN (1999). *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. John Wiley & Sons.
- A. HAASE (1990). Snapshot FLASH MRI: applications to T1, T2, and chemical-shift imaging. *Magnetic Resonance in Medicine*, 13(1):77–89.

- A. HAASE, J. FRAHM, D. MATTHAEI, W. HANICKE, E K.-D. MERBOLDT (1986). FLASH imaging. Rapid NMR imaging using low flip-angle pulses. *Journal of Magnetic Resonance (1969)*, 67(2):258–266.
- A. HAASE, D. MATTHAEI, R. BARTKOWSKI, E. DÜHMKE, E D. LEIBFRITZ (1989). Inversion recovery snapshot FLASH MR imaging. *Journal of computer assisted tomography*, 13(6):1036–1040.
- C. HAGEDORN, M. I. PROCTOR, E L. GOLDSTEIN (2011). Automatic Analysis of Singleton and Geminate Consonant Articulation using Real-time Magnetic Resonance Imaging. In *Proceedings of Interspeech, Florence, Italy*.
- C. HAGEDORN, M. I. PROCTOR, L. GOLDSTEIN, M. L. G. TEMPINI, E S. S. NARAYANAN (2012). Characterizing Covert Articulation in Apraxic Speech Using real-time MRI. In *Proceedings of Interspeech, Portland, USA*.
- L. HAHN E L. QUEDNAU (2007). A lateral pós-vocálica no português de Londrina: análise variacionista e estrutura silábica. *Letras de Hoje*, 42(3):100–113.
- G. HAMARNEH E T. GUSTAVSSON (2000). Combining snakes and active shape models for segmenting the human left ventricle in echocardiographic images. In *Proceedings of IEEE Computers in Cardiology*, volume 27, páginas 115–118.
- G. HAMARNEH, J. YANG, C. MCINTOSH, E M. LANGILLE (2005). 3D live-wire based semi-automatic segmentation of medical images. In *Proceedings of SPIE Medical Imaging*, volume 5747, páginas 1597–1603.
- R. H. HASHEMI, W. G. BRADLEY, E C. J. LISANTI (2012). *MRI: the basics*. Lippincott Williams & Wilkins.
- K. HEBERLEIN E X. HU (2006). Auto-calibrated parallel spiral imaging. *Magnetic Resonance in Medicine*, 55(3):619–625.
URL <http://dx.doi.org/10.1002/mrm.20811>
- R. HEIDEMANN, O. OZSARLAK, P. PARIZEL, J. MICHIELS, B. KIEFER, V. JELLUS, M. MÜLLER, F. BREUER, M. BLAIMER, M. A. GRISWOLD, E P. M. JAKOB (2003). A brief review of parallel magnetic resonance imaging. *European Radiology*, 13(10):2323–2337.
- J. HENNIG, M. WEIGEL, E K. SCHEFFLER (2003). Multiecho sequences with variable refocusing flip angles: optimization of signal behavior using smooth transitions between pseudo steady states (TRAPS). *Magnetic resonance in medicine*, 49(3):527–535.
- S. HO, H. CODY, E G. GERIG (2003). SNAP: A software package for User-Guided Geodesic Snake Segmentation. Relatório técnico, University of North Carolina, USA.
- K. HONDA (2002). Evolution of vowel production studies and observation techniques. *Acoustical Science and Technology*, 23(4):189.

- M. HONDA (2003). Human speech production mechanisms. *NTT Technical Review*, 1(2):24–29.
- P. HOOLE E N. NGUYEN (1999). Electromagnetic articulography. In W. HARDCASTLE E N. HEWLETT (editores), *Coarticulation: Theory, Data and Techniques*, páginas 260–269. Cambridge University Press, Cambridge.
- P. HOOLE, A. WISMUELLER, G. LEINSINGER, C. KROOS, A. GEUMANN, E M. INOUE (2000). Analysis of tongue configuration in multi-speaker, multi-volume MRI data. In *Proceedings of the 5th International Seminar on Speech Production (ISSP), Kloster Seeon, Bavaria, Germany*, páginas 157–160.
- P. HOOLE, A. ZIERDT, E C. GENG (2003). Beyond 2D in articulatory data acquisition and analysis. In *Proceedings of the 15th International Conference of Phonetic Sciences (ICPhS) - Barcelona, Spain*, páginas 265–268.
- D. HORA (2006). Vocalização da lateral /l/: correlação entre restrições sociais e estruturais. *SCRIPTA*, 9(18):29–44.
- HPA (2008). Protection of patients and volunteers undergoing MRI procedures. Relatório técnico, Health Protection Agency - RCE7- Advice of Health Protection Agency - UK.
- F. HUANG, J. AKAO, S. VIJAYAKUMAR, G. R. DUENSING, E M. LIMKEMAN (2005). k-t GRAPPA: A k-space implementation for dynamic MRI with high reduction factor. *Magnetic Resonance in Medicine*, 54(5):1172–1184.
URL <http://dx.doi.org/10.1002/mrm.20641>
- T. HUEBER, G. CHOLLET, B. DENBY, E M. STONE (2008). Acquisition of ultrasound, video and acoustic speech data for a silent-speech interface application. In *Proceedings of the 8th International Seminar on Speech Production (ISSP), Strasbourg, France*, páginas 365–369.
- S. HUETTEL, A. SONG, E G. MCCARTHY (2004). *Functional Magnetic Resonance*, volume 1. Sinauer Associates Inc, Massachusetts.
- ICNIRP (2004). Medical magnetic resonance MR procedures: protection of patients. Relatório Técnico 2, International Commission on Non-Ionizing Radiation Protection.
- D. IDIYATULLIN, C. CORUM, S. MOELLER, H. S. PRASAD, M. GARWOOD, E D. R. NIXDORF (2011). Dental magnetic resonance imaging: making the invisible visible. *Journal of endodontics*, 37(6):745–752.
- D. IDIYATULLIN, C. A. CORUM, D. R. NIXDORF, E M. GARWOOD (2013). Intraoral approach for imaging teeth using the transverse B1 field components of an occlusally oriented loop coil. *Magnetic Resonance in Medicine*.
- IEC (2002). Medical electrical equipment– part 2–33: Particular requirements for the safety of magnetic resonance equipment for medical diagnosis. Relatório técnico, International Electrotechnical Commission - Genève.

- K. ISKAROUS, L. GOLDSTEIN, D. H. WHALEN, M. K. TIEDE, E P. E. RUBIN (2003). CASY: The Haskins configurable articulatory synthesizer. In *International Congress of Phonetic Sciences (ICPhS), Barcelona, Spain*, volume 1, páginas 185–188.
- A. ISRAEL, M. I. PROCTOR, L. GOLDSTEIN, K. ISKAROUS, E S. S. NARAYANAN (2012). Emphatic segments and emphasis spread in lebanese arabic: a real-time magnetic resonance imaging study. In *Proceedings of InterSpeech*.
- L. JESUS E C. SHADLE (2005). Acoustic analysis of European Portuguese uvular [X,R] and voiceless tapped alveolar [r] fricatives. *Journal of the International Phonetic Association*, 35(1):27–44.
- H. JUNG, K. SUNG, K. S. NAYAK, E. Y. KIM, E J. C. YE (2009). k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI. *Magnetic Resonance in Medicine*, 61(1):103–116.
- S. KAJI, P. C. YANG, A. B. KERR, W. W. TANG, C. H. MEYER, A. MACOVSKI, J. M. PAULY, D. G. NISHIMURA, E B. S. HU (2001). Rapid evaluation of left ventricular volume and mass without breath-holding using real-time interactive cardiac magnetic resonance imaging system. *Journal of the American College of Cardiology*, 38(2):527–533.
- E. KANAL, A. J. BARKOVICH, C. BELL, J. P. BORGSTEDE, W. G. BRADLEY, J. W. FROELICH, T. GILK, J. R. GIMBEL, J. GOSBEE, E. KUHN-KAMINSKI, *et al.* (2007). ACR guidance document for safe MR practices: 2007. *American Journal of Roentgenology*, 188(6):1447–1474.
- A. A. KANE, J. A. BUTMAN, R. MULLICK, M. SKOPEC, E P. CHOYKE (2002). A new method for the study of velopharyngeal function using gated magnetic resonance imaging. *Plastic and reconstructive surgery*, 109(2):472–481.
- H. KANG (2005). G-Wire: A livewire segmentation algorithm based on a generalized graph formulation. *Pattern Recognition Letters*, 22(26):2042–2051.
- S. KARACAY, E. AKIN, K. ORTAKOGLU, E A. O. BENGI (2006). Dynamic MRI evaluation of tongue posture and deglutitive movements in a surgically corrected open bite. *The Angle Orthodontist*, 76(6):1057–1065.
- M. KASS, A. P. WITKIN, E D. TERZOPOULOS (1987). Snakes: Active Contour Models. *International Journal of Computer Vision*, 1(4):321–331.
- P. KEATING (1988). Palatals as complex segments: X-ray evidence. *UCLA Working Papers on Phonetics*, 69:77–89.
- G. KEDROVA, N. ANISIMOV, E J. PIROGOV (2009). MRI-based contrastive study of nasal and oral labial consonants articulations in Russian. In *Proceedings of the International Workshop - Nasal 2009 - Montpellier, France*.
- G. KEDROVA, N. ANISIMOV, L. ZAHAROV, E J. PIROGOV (2006). Contrastive study of the MRI representation of Russian vowel articulation (against French, German and Korean analogues). In *33rd International Acoustical Conference - EAA Symposium ACOUSTICS-Slovakia*.

- N. KEDROVA, V. ANISIMOV, E. A. PIROGOV (2008). On-line visualization of speech organs using MRI: a 3D approach to speech articulation modeling. In *Proceedings of the 11th IASTED International Conference on Computers and Advanced Technology in Electronics*.
- D. KIM, H. A. DYVORNE, R. OTAZO, L. FENG, D. K. SODICKSON, E. V. S. LEE (2012). Accelerated phase-contrast cine MRI using k-t SPARSE-SENSE. *Magnetic Resonance in Medicine*, 67(4):1054–1064.
URL <http://dx.doi.org/10.1002/mrm.23088>
- H. KIM (2004). Stroboscopic-Cine MRI data on Korean coronal plosives and affricates: Implications for their place of articulation as alveolar. *Phonetica*, 61(4):234–251.
- J. KIM, A. C. LAMMERT, P. KUMAR GHOSH, E. S. S. NARAYANAN (2014). Co-registration of speech production datasets from electromagnetic articulography and real-time magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 135(2):EL115–EL121.
URL <http://scitation.aip.org/content/asa/journal/jasa/135/2/10.1121/1.4862880>
- Y.-C. KIM (2010). *Fast Upper Airway MRI of Speech*. Tese de Doutorado, University of Southern California, California, USA.
- Y.-C. KIM, R. M. LEBEL, Z. WU, S. L. D. WARD, M. C. KHOO, E. K. S. NAYAK (2013). Real-time 3D Magnetic Resonance Imaging of the pharyngeal airway in sleep apnea [available online]. *Magnetic Resonance in Medicine*.
- Y.-C. KIM, S. NARAYANAN, E. K. NAYAK (2009). Accelerated three-dimensional upper airway MRI using Compressed Sensing. *Magnetic Resonance in Medicine*, 61:1434–1440.
- Y.-C. KIM, S. NARAYANAN, E. K. NAYAK (2011). Visualization of vocal tract shape using interleaved real-time MRI of multiple scan planes. In *Proceedings of Interspeech, Florence, Italy*.
- Y.-C. KIM, S. S. NARAYANAN, E. K. S. NAYAK (2010). Improved Real-time MRI of Oral-Velar Coordination Using a Golden-ratio Spiral View Order. In *Proceedings of Interspeech, Makuhari, Japan*.
- Y.-C. KIM, M. PROCTOR, S. NARAYANAN, E. K. NAYAK (2012). Improved imaging of lingual articulation using real-time multislice MRI. *Journal of Magnetic Resonance Imaging*, 35(4):943–948.
- R. A. KRAKOW (1993). Nonsegmental influences on velum movement patterns: syllables, sentences, stress, and speaking rate. *Nasals, nasalization, and the velum*, 5:87–118.
- P. D. KROEBER (1999). *The Salish language family: Reconstructing syntax*. University of Nebraska Press.
- B. KRÖGER, J. GOTTO, S. ALBERT, E. C. NEUSCHAEFER-RUBE (2005). A visual articulatory model and its application to therapy of speech disorders: a pilot study. *Speech production and perception: Experimental analyses and models. ZAS Papers in Linguistics*, 40:79–94.

- B. KRÖGER, R. WINKLER, C. MOOSHAMMER, E B. POMPINO-MARSCHALL (2000). Estimation of vocal tract area function from Magnetic Resonance Imaging: Preliminary results. In *Proceedings of the 5th International Seminar on Speech Production (ISSP)- Kloster Seeon, Bavaria, Germany*, páginas 333–336.
- A. LACERDA E B. F. HEAD (1966). Análise de sons nasais e nasalizados do português. *Revista do Laboratório de Fonética Experimental*, 6:5–43.
- P. LADEFOGED (1988). Hierarchical Features of the International Phonetic Alphabet. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, volume 14, páginas 124–141.
- P. LADEFOGED (2001). *Vowels and Consonants: An introduction to the sounds of languages*. Blackwell Publishers, Inc, Oxford.
- P. LADEFOGED E I. MADDIESON (1986). Some of the sounds of the world languages: preliminary version. *UCLA Working Papers in Phonetics*, 64.
- J. LAFUENTE (2000). *Atlas de Tecnología de la Resonancia Magnética*. Mallinckrodt publications, Madrid.
- A. LAMMERT, E. BRESCH, D. BYRD, L. GOLDSTEIN, E S. S. NARAYANAN (2009). An articulatory study of lexicalized and epenthetic schwa using real time magnetic resonance imaging. In *Proceedings of the Meeting of the Acoustical Society of America*. Portland, Oregon.
- A. LAMMERT, E. BRESCH, L. GOLDSTEIN, E S. S. NARAYANAN (2010a). Gestural control in the English past-tense suffix: an articulatory study using real time MRI. In *Proceedings of LabPhon, 2010*.
- A. LAMMERT, M. I. PROCTOR, E S. S. NARAYANAN (2010b). Data-Driven Analysis of Realtime Vocal Tract MRI using Correlated Image Regions. In *Proceedings of Interspeech, Makuhari, Japan*.
- D. LARKMAN (2009). Tools to speed up MR imaging. In *Proceedings of the ESMRMB Congress, Antalya*.
- J. LAVER (1980). *The Phonetic Description of Voice Quality*. Cambridge studies in linguistics ; 31. Cambridge University Press, Cambridge, primeira edição.
- S. LEE, E. BRESCH, J. ADAMS, A. KAZEMZADEH, E S. NARAYANAN (2006). A study of emotional speech articulation using a fast magnetic resonance imaging technique. In *Proceedings of Interspeech, Pittsburgh, Pennsylvania, USA*.
- S. LEE E S. S. NARAYANAN (2010). Vocal tract contour analysis of emotional speech by the functional data curve representation. In *Proceedings of Interspeech, Makuhari, Japan*.
- J. LIANG, T. MCINERNEY, E D. TERZOPOULOS (2006). United snakes. *Medical Image Analysis*, 10:215–233.
- Z.-P. LIANG E P. LAUTERBUR (2000). *Principles of Magnetic Resonance Imaging- A signal processing perspective*. IEEE Press, New York.

- E. LIN (2012). Body MRI sequences: A conceptual framework. *Applied Radiology*, 41(1):16.
- Z. LIN, J. JIN, E H. TALBOT (2000). Unseeded region growing for 3D image segmentation. In *Selected papers from the Pan-Sydney workshop on Visualisation-Volume 2*, páginas 31–37. Australian Computer Society, Inc.
- J. LIU, J. K. UDUPA, D. ODHNERA, J. M. McDONOUGH, E R. ARENS (2003). System for upper airway segmentation and measurement with MR imaging and fuzzy connectedness. *Academic radiology*, 10(1):13–24.
- D. LOPES, J. MARTINS, J. CAMPOS, E E. PIRES (2007). Modelação geométrica de Estruturas Humanas baseada em imagens de Tomografia Computorizada. In *Proceedings do Congresso de Métodos Numéricos em Engenharia /CILANCE, Porto, Portugal*.
- W. E. LORENSEN E H. E. CLINE (1987). Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21:163–169.
URL <http://doi.acm.org/10.1145/37402.37422>
- M. LOUSADA, A. MENDES, A. VALENTE, E A. HALL (2012). Standardization of a Phonetic-Phonological Test for European-Portuguese Children. *Folia Phoniatrica et Logopaedica*, 64(3):151–156.
- L. LOVATTO, A. AMELOT, L. CREVIER-BUCHMAN, P. BASSET, J. VAISSIÈRE, *et al.* (2007). A fiberscopic analysis of nasal vowels in Brazilian Portuguese. In *Proceedings of the 16th International Congress of Phonetic Sciences(ICPhS)- Saarbrücken, Germany*, páginas 549–552.
- M. LUSTIG, D. DONOHO, E J. M. PAULY (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195.
URL <http://dx.doi.org/10.1002/mrm.21391>
- M. LUSTIG, D. L. DONOHO, J. M. SANTOS, E J. M. PAULY (2008). Compressed sensing MRI: A look at how CS can improve on current imaging techniques. *IEEE Signal Processing Magazine*, páginas 72–82.
- I. MADDIESON (2007). Areal distribution of nasalized vowels. In *Proceedings of the 16th International Congress of Phonetic Sciences(ICPhS)- Saarbrücken, Germany*.
- R. MALLADI, J. A. SETHIAN, E B. C. VEMURI (1995). Shape Modeling with Front Propagation: A Level Set Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175.
- P. MANSFIELD (1977). Multi-planar image formation using NMR spin echoes. *J. Phys. C: Solid State Phys.*, 10(3):L55–L58.
- MANUAL MEVISLAB (2010). *MeVisLab reference manual*. MeVis Medical Solutions 2003–2013, Alemanha.

- I. MARQUES (2010). *Variação fonética da lateral alveolar no Português Europeu*. Tese de Mestrado, Universidade de Aveiro, Aveiro.
- M. R. D. MARTINS (1988). *Ouvir falar - Introdução à Fonética do Português*. Universitária - Linguística. Editorial Caminho, Lisboa, segunda edição.
- P. MARTINS (2007). *Ressonância Magnética no Estudo da Produção do Português Europeu*. Tese de Mestrado, Universidade de Aveiro, Aveiro.
- P. MARTINS, I. CARBONE, A. PINTO, A. SILVA, E A. TEIXEIRA (2008). European Portuguese MRI based speech production studies. *Speech Communication*, 50(11-12):925–952.
- P. MARTINS, I. CARBONE, A. SILVA, E A. TEIXEIRA (2007). An MRI Study of European Portuguese Nasals. In *Proceedings of Interspeech, Antwerp, Belgium*.
- P. MARTINS, I. CARBONE, A. SILVA, E A. TEIXEIRA (2011a). Effects coarticulatoires sur le portugais européen: une première étude IRM. In M. EMBARKI E C. DODANE (editores), *La coarticulation: Des indices à la représentation*, Langue & Parole, páginas 101–115. Editions L'Harmattan, Paris.
- P. MARTINS, C. OLIVEIRA, C. FERREIRA, A. SILVA, E A. TEIXEIRA (2011b). 3D MRI and Semi-automatic techniques applied to the study of European Portuguese lateral sounds. In *9th International Seminar on Speech Production (ISSP), Montreal, Canadá*, páginas 235–242.
- P. MARTINS, C. OLIVEIRA, A. SILVA, E A. TEIXEIRA (2010). Articulatory characteristics of European Portuguese Laterals: An 2D and 3D MRI Study. In *FALA 2010, Vigo, Spain*.
- P. MARTINS, C. OLIVEIRA, S. SILVA, C. FERREIRA, A. SILVA, E A. TEIXEIRA (2012a). Polygonal Mesh Comparison applied to the study of European Portuguese Sounds. *International Journal of Creative Interfaces and Computer Graphics*, 3:28–44.
- P. MARTINS, C. OLIVEIRA, S. SILVA, A. SILVA, E A. TEIXEIRA (2011c). Tongue segmentation from MRI images using ITK-SNAP: Preliminary Evaluation. In *International Conferences Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP) 2011, Rome, Italy*, páginas 3–10.
- P. MARTINS, C. OLIVEIRA, S. SILVA, E A. TEIXEIRA (2012b). Velar movement in European Portuguese nasal vowels. In *Proceedings of IberSPEECH 2012 - VII Jornadas de Tecnologia dela Habla and III Iberian SLtech Workshop, Madrid, Spain*.
- P. MARTINS, S. SILVA, C. OLIVEIRA, C. FERREIRA, A. SILVA, E A. TEIXEIRA (2011d). Investigating the differences between European Portuguese sounds: An Approach using Polygonal Mesh Comparison. In *Proceedings of the V Ibero-American Symposium in Computer Graphics SIACG-2011, Faro, Portugal*, páginas 209–216.
- S. MASAKI, Y. NOTA, S. TAKANO, H. TAKEMOTO, T. KITAMURA, E K. HONDA (2008). Integrated magnetic resonance imaging methods for speech science and technology. *Journal of the Acoustical Society of America*, 123(5):3734.

- S. MASAKI, M. TIEDE, K. HONDA, Y. SHIMADA, I. FUJIMOTO, Y. NAKAMUR, E N. NINOMIYA (1997). MRI observation of dynamic articulatory movements using a synchronizd sampling method. *Journal of the Acoustical Society of America (JASA)*, 102(5:2):3166.
- S. MASTER, P. PONTES, E M. BEHLAU (1991). Configuração do trato vocal na emissão das vogais nasais do português brasileiro. *Acta AWHO*, 2(10):67–71.
- M. H. M. MATEUS (1975). *Aspectos de Fonologia Portuguesa*. Centro de Estudos Filológicos, Lisboa.
- M. H. M. MATEUS E E. D'ANDRADE (2000). *Phonology of Portuguese*. Oxford University Press, Oxford.
- M. H. M. MATEUS, I. FALÉ, E M. J. FREITAS (2005). *Fonética e Fonologia do Português*. Universidade Aberta, Lisboa.
- K. MATHIAK, I. HERTRICH, W. E. KINCSES, U. KLOSE, H. ACKERMANN, E W. GRODD (2000). Stroboscopic articulography using fast magnetic resonance imaging. *International journal of language & communication disorders/Royal College of Speech & Language Therapists*, 35(3):419–25.
- M. MATTA MACHADO (1993). Fenômenos de nasalização Vocálica em Português. Estudo cine-radiográfico. *Cadernos de Estudos Linguísticos*, 25:113–117.
- T. MCINERNEY E D. TERZOPOULOS (1996). Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108.
- K. MÁDY, R. SADER, A. ZIMMERMANN, P. HOOLE, A. J. BEER, E H. ZEILHOFER (2002). Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction. In *4th International Speech Motor Conference*, páginas 142–145.
- B. MEDEIROS E D. DEMOLIN (2006). Vogais nasais do Português Brasileiro: Um Estudo IRM. *Revista da ABRALIN*, 5:131–142.
- B. R. MEDEIROS (2011). Nasal coda and vowel nasality in brazilian portuguese. In *Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology, Somerville*, páginas 33–45.
- P. MERMELSTEIN (1973). Articulatory model for the study of speech production. *Journal of The Acoustic Society of America (JASA)*, 53(4):1070–1082.
- MHRA (2007). Safety guidelines for magnetic resonance imaging equipment in clinical use. Relatório técnico, Medicines and Healthcare products Regulatory Agency (MHRA)- Device bulletin - UK.
- H. J. MICHAELY, H. KRAMER, N. OESINGMANN, K.-P. LODEMANN, M. F. REISER, E S. O. SCHONBERG (2007). Semiquantitative assessment of first-pass renal perfusion at 1.5 T: comparison of 2D saturation recovery sequences with and without parallel imaging. *American Journal of Roentgenology*, 188(4):919–926.

- G. MODHA, B. BERNHARDT, R. CHURCH, E P. BACSFALVI (2008). Case study using ultrasound to treat /r/. *International Journal of Language & Communication Disorders*, 43(3):323–329.
- M. A.-J. MOHAMMAD (1999). *Dynamic measurements of speech articulators using MRI*. Tese de doutoramento, University of Southampton.
- K. MOLL (1960). Cinefluorographic techniques in speech research. *Journal of Speech and Hearing Research*, (3):227–241.
- K. L. MOLL E R. G. DANILOFF (1971). Investigation of the timing of velar movements during speech. *The Journal of the Acoustical Society of America*, 50(2B):678–684.
- J. MONTAGU (2004). Les sons sous-jacents aux voyelles nasales an français parisien: indices perceptifs des changement. In *Proceedings of the XXV èmes Journées d' Étude sur la Parole (JEP)*, Maroc, páginas 385–388.
- D. MONTEIRO (2012). *Variação dialectal das laterais do Português Europeu*. Tese de Mestrado, Universidade de Aveiro, Aveiro.
- C. MOOSHAMMER, P. PERRIER, S. FUCHS, C. GENG, E D. PAPE (2004). An EMMA and EPG study on token-to-token variability. *AIPUK*, 36:47–63.
- J. MORAES E W. L. WETZELS (1992). Sobre a duração dos segmentos vocálicos nasais e nasalizados em português. Um exercício de fonologia experimental. *Cadernos de Estudos Linguísticos*, 23:153–166.
- J. N. MORELLI, V. M. RUNGE, F. AI, U. ATTENBERGER, L. VU, S. H. SCHMEETS, W. R. NITZ, E J. E. KIRSCH (2011). An image-based approach to understanding the physics of MR artifacts. *Radiographics*, 31(3):849–866.
- E. N. MORTENSEN E W. A. BARRETT (1995). Intelligent scissors for image composition. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '95, páginas 191–198. ACM, New York, NY, USA.
- L. MOUTINHO (2000). *Uma Introdução ao Estudo da Fonética e Fonologia do Português*. Plátano - Edições Técnicas.
- M. NAKAMURA (2009). Exploring Vocalization of /l/ in English: an EPG and EMA study. In *Proceedings of Interspeech, Brighton, UK*.
- S. NARAYANAN, A. ALWAN, E K. HAKER (1994). Three-dimensional tongue shapes of sibilant fricatives. *Journal of the Acoustical Society of America*, 96(5, Pt. 2):3342.
- S. NARAYANAN, A. ALWAN, E K. HAKER (1995). An Articulatory Study of Fricative Consonants using MRI. *Journal of the Acoustical Society of America (JASA)*, 98(3):1325–1347.
- S. NARAYANAN, A. ALWAN, E K. HAKER (1997a). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals. *Journal of the Acoustical Society of America (JASA)*, 101(2):1064–1077.

- S. NARAYANAN, A. ALWAN, E Y. SONG (1997b). New Results in Vowel Production: MRI, EPG, and Acoustic Data. In *Proceedings of Eurospeech '97, Rhodes, Greece.*, páginas 1007–1010.
- S. NARAYANAN, D. BIRD, E A. KAUN (1999). Geometry, kinematics and acoustics of Tamil liquid consonants. *Journal Acoustical Society of America*, 106:1993–2007.
- S. NARAYANAN, E. BRESCH, P. K. GHOSH, L. GOLDSTEIN, A. KATSAMANIS, Y.-C. KIM, A. LAMMERT, M. I. PROCTOR, V. RAMANARAYANAN, E Y. ZHU (2011). A Multimodal Real-Time MRI Articulatory Corpus for Speech Research. In *Proceedings of Interspeech, Florence, Italy*.
- S. NARAYANAN, A. KAUN, D. BYRD, P. LADEFOGED, E A. ALWAN (1996). Liquids in tamil. In *4th International Conference on Spoken Language Processing (ICSLP)*.
- S. NARAYANAN, K. NAYAK, S. LEE, A. SETHY, E D. BYRD (2004). An approach to real-time magnetic resonance imaging for speech production. *Journal of Acoustical Society of America*, 115(4):1771–1776.
- K. NAYAK, B. HARGREAVES, B. HU, D. NISHIMURA, J. PAULY, E C. MEYER (2005). Spiral Balanced Steady-State Free Precession Cardiac Imaging. *Magnetic Resonance in Medicine*, 53:1468–1473.
- K. NAYAK E B. HU (2005). The future of real-time cardiac magnetic resonance imaging. *Current cardiology reports*, 7(1):45–51.
- M. NESSAIVER, Y. KAHANA, A. PARITSKY, M. STONE, V. PARTHASARATHY, E A. PARITSKY (2006). Recording high quality speech during tagged cine-MRI studies using a fiber optic microphone. *Journal of Magnetic Resonance Imaging*, 23(1):92–7.
- NGUYEN (2006). Using ITK-SNAP to Segment Out Lateral Ventricles. Relatório técnico, UC Davis Medical Center - The MIND Institute.
- A. NIEBERGALL, S. ZHANG, E. KUNAY, G. KEYDANA, M. JOB, M. UECKER, E J. FRAHM (2012). Real-time MRI of speaking at a resolution of 33 ms: Undersampled radial FLASH with nonlinear inverse reconstruction. *Magnetic Resonance in Medicine*, 69:477–485.
URL <http://dx.doi.org/10.1002/mrm.24276>
- W. J. NIESSEN, B. ROMENY, E M. VIERGEVER (1998). Geodesic Deformable Models for Medical Image Analysis. *IEEE Transactions on Medical Imaging*, 17(4):634–641.
- H. NISHIMOTO, Y. SHIMADA, I. FUJIMOTO, S. MASAKI, E K. MURASE (2011). Observation of the Tongue Movement Producing Alveolar Trill by High Frame Rate MRI-Movie. In *Proceedings of the 9th International Seminar on Speech Production (ISSP), Montreal, Canadá*, páginas 33–40.
- M. OKUBO, M. SUZUKI, A. HORIUCHI, S. OKABE, K. IKEDA, S. HIGANO, H. MITANI, W. HIDA, T. KOBAYASHI, E J. SUGAWARA (2006). Morphologic analyses of mandible and upper airway soft tissue by MRI of patients with obstructive sleep apnea hypopnea syndrome. *SLEEP*, 29(7):909–915.

- C. OLIVEIRA (2009). *Do Grafema ao Gesto: contributos linguísticos para um sistema de síntese de base articulatória*. Tese de doutoramento, Universidade de Aveiro, Aveiro.
- C. OLIVEIRA, P. MARTINS, S. S. SILVA, E A. J. TEIXEIRA (2012). An MRI study of the oral articulation of European Portuguese nasal vowels. In *Proceedings of Interspeech, Portland, USA*.
- C. OLIVEIRA, P. MARTINS, I. MARQUES, P. SÁ-COUTO, E A. TEIXEIRA (2011). An articulatory and acoustic study of european portuguese /l/. In *International Congress on Phonetics Science, Hong-Kong, China*.
- C. OLIVEIRA E A. TEIXEIRA (2007a). Nova Base de Dados EMMA relativa às Nasais e Laterais do Português Europeu. Relatório técnico, IEETA, Universidade de Aveiro, Aveiro.
- C. OLIVEIRA E A. TEIXEIRA (2007b). On Gestures timing in European Portuguese nasals. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)- Saarbrücken, Germany*.
- C. OLIVEIRA, A. TEIXEIRA, E P. MARTINS (2009). Speech Rate Effects on European Portuguese Nasal Vowels. In *Proceedings of the 10th Annual Conference of ISCA Interspeech - Brighton, UK*.
- C. OLIVEIRA, A. TEIXEIRA, E P. MARTINS (2010). Towards an articulatory characterization of european portuguese /l/. In *Proceedings of ExLing 2010, Atenas, Grécia*, páginas 133–137.
- L. OLIVEIRA E S. MARIN (2005). Patterns of velum coordination in Brazilian Portuguese. In *Proceedings of Phonetics and Phonology in Iberia Conference - Barcelona, Spain*.
- D. ONG E M. STONE (1998). Three-dimensional vocal tract shapes in /r/ and /l/: A study of MRI, ultrasound, electropalatography, and acoustics. *Phonoscope*, 1(1):1–13.
- OPTOACOUSTICS (2010). *FOMRI III Dual Channel Microphone System for fMRI-User guide*. Optoacoustics Ltd - Sound Solutions From Light Technology, Or Yehuda 60212 ISRAEL.
- S. OSHER E N. PARAGIOS (2003). *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer Verlag. ISBN 0-387-95488-0.
- S. OSHER E J. SETHIAN (1988). Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi Formulation. *Computer Physics*, 22(79):12–49.
- J. S. PERKELL (1971). Physiology of speech production: Results and implications of a quantitative cineradiographic study. *Language*, 47(1):237–244.
- P. PERRIER, Y. PAYAN, M. ZANDIPOUR, E J. PERKELL (2003). Influences of tongue biomechanics on velar-stop consonants: A modeling study. *Journal of Acoustical Society of America*, 103:1615–1627.

- J. L. PERRY, D. P. KUEHN, B. P. SUTTON, M. S. GOLDWASSER, E A. D. JEREZ (2011). Craniometric and velopharyngeal assessment of infants with and without cleft palate. *Journal of Craniofacial Surgery*, 22(2):499–503.
- J. L. PERRY, B. P. SUTTON, D. P. KUEHN, E J. K. GAMAGE (2013). Using MRI for Assessing Velopharyngeal Structures and Function [available online]. *The Cleft Palate-Craniofacial Journal*.
- D. PHAM, X. CHENYANG, E J. PRINCE (2000). Current Methods in Medical Imaging Segmentation. *Annu. Rev. Biomed. Eng.*, 2:315–337.
- J. PICADO (2010). Estruturas Discretas: textos de apoio. Capítulo 2. Relatório técnico, Universidade de Coimbra, Coimbra.
- A. PINHO E F. MARGOTTI (2010). A Variação da lateral pósvoçálica /l/ no Português do Brasil. *Working papers em Linguística*, 2:67–88.
- R. POHMANN, G. SHAJAN, J. HOFFMANN, G. BUDDE, J. HAGBERG, O. BIERI, S. BISDAS, U. ERNEMANN, M. WEIGEL, J. HENNIG, G. CHADZYNSKI, E K. SCHEFFLER (2013). Imaging and spectroscopy at 9.4 Tesla: First results on patients and volunteers. *MagnetonFLASH - ISMRM edition*, 52(2):58–67.
- M. POON, G. HAMARNEH, E R. ABUGHARBIEH (2007). Segmentation of complex objects with non-spherical topologies from volumetric medical images using 3D livewire. In *Proceedings of SPIE - Medical Imaging*, volume 6512-31, páginas 1–10.
- M. POON, G. HAMARNEH, E R. ABUGHARBIEH (2008). Efficient interactive 3D Livewire segmentation of complex objects with arbitrary topology. *Computerized Medical Imaging and Graphics*, 32(8):639–650.
- A. PRAHLER (1998). *Analysis and synthesis of the American English lateral consonant*. Tese de Doutoramento, Massachusetts Institute of Technology, Cambridge, Massachusetts EUA.
- W. PRATT (1978). *Digital Image Processing*. John Wiley e Sons.
- M. PROCTOR (2009). *Gestural characterization of a phonological class: The liquids*. Tese de Doutoramento, Yale University, USA.
- M. PROCTOR, L. GOLDSTEIN, A. LAMMERT, D. BYRD, A. TOUTIOS, E S. NARAYANAN (2013). Velic Coordination in French Nasals: a Real-time Magnetic Resonance Imaging Study. In *Proceedings of Interspeech, Lyon, France*.
- M. PROCTOR, L. H. LU, Y. ZHU, L. GOLDSTEIN, E S. NARAYANAN (2012). Articulation of Mandarin Sibilants: a multi-plane realtime MRI study. In *Proceedings of the International Conference on Speech Science and Technology, Sydney, Australia*.
- M. PROCTOR, S. NARAYANAN, E K. NAYAK (2010a). Para-linguistic mechanisms of production in Human 'Beatboxing': a Real-Time Magnetic Resonance Imaging study. In *Proceedings of InterSinging, Tokyo, Japan*.

- M. I. PROCTOR, A. LAMMERT, A. KATSAMANIS, L. GOLDSTEIN, C. HAGEDORN, E S. NARAYANAN (2011). Direct Estimation of Articulatory Kinematics from Real-time Magnetic Resonance Image Sequences. In *Proceedings of Interspeech, Florence, Italy*.
- M. I. PROCTOR, C. H. SHADLE, E K. ISKAROUS (2010b). Pharyngeal articulation in the production of voiced and voiceless fricatives. *The Journal of the Acoustical Society of America*, 127:1507.
- K. PRUESSMANN, M. WEIGER, M. SCHEIDEGGER, E P. BOESIGER (1999). SENSE: Sensitivity Encoding for Fast MRI. *Magnetic Resonance in Medicine*, 42:952–956.
- K. P. PRUESSMANN (2006). Encoding and reconstruction in parallel MRI. *NMR in Biomedicine*, 19(3):288–299.
URL <http://dx.doi.org/10.1002/nbm.1042>
- T. PRUTHI, C. Y. ESPY-WILSON, E B. H. STORY (2007). Simulation and analysis of nasalized vowels based on magnetic resonance imaging data. *The Journal of the Acoustical Society of America*, 121:3858.
- V. RAMANARAYANAN, a. L. G. DANI BYRD, E S. S. NARAYANAN (2010). A joint acoustic-articulatory study of nasal spectral reduction in read versus spontaneous speaking styles. In *Proceedings of the Speech Prosody Conference, Chicago, Illinois, USA*.
- V. RAMANARAYANAN, A. LAMMERT, D. BYRD, L. GOLDSTEIN, E S. NARAYANA (2011). Planning and execution in soprano singing and speaking behavior: An acoustic-articulatory study using real-time MR. In *Proceedings of the International Seminar on Speech Production (ISSP'11), Montreal, Canada*.
- D. RECASENS (1990). The articulatory characteristics of palatal consonants. *Journal of Phonetics*, 18:267–280.
- D. RECASENS (2004). Darkness in [ɫ] as a scalar phonetic property: implications for phonology and articulatory control. *Clinical Linguistics & Phonetics*, 18(6-8):593–603.
- D. RECASENS (2012). A cross-language acoustic study of initial and final allophones of /l/. *Speech Communication*, 54:368–383.
- D. RECASENS E A. ESPINOSA (2005). Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: evidence from two Catalan dialects. *Journal of the International Phonetic Association*, 35(1):1–25.
- D. RECASENS E A. ESPINOSA (2006). Articulatory, positional and contextual characteristics of palatal consonants: Evidence from Majorcan Catalan. *Journal of Phonetics*, 34:295–318.
- D. RECASENS E A. ESPINOSA (2009). Lingual kinematics and coarticulation for alveolopalatal and velar consonants in Catalan. *Journal of Acoustical Society of America*, 125:2288–2298.
- D. RECASENS, E. FARNETANI, J. FONTDEVILA, E M. PALLARÉS (1993). An Electropalatographic study of alveolar and palatal consonants in Catalan and Italian. *Language and Speech*, 36:213–234.

- D. RECASENS, J. FONTDEVILA, E M. D. PALLARÉS (1996). Linguopalatal coarticulation and alveolar-palatal correlations for velarized and non-velarized /l/. *Journal of Phonetics*, 24:165–185.
- P. REENEN (1982). *Phonetic Feature Definitions: Their integration into phonology and their relation to speech; a case study of the feature NASAL*. De Gruyter Mouton.
- C. REIS E R. ESPESER (2006). Estudo electropalatográfico de fones consonantais e vocálicos do português brasileiro. *Estudos da Língua(gem): Questões de Fonética e Fonologia -Homenagem a Luis Carlos Cagliari*, 3:181–204.
- N. RIBEIRO, D. FERNANDES, D. LOPES, J. FOLGADO, E P. FERNANDES (2009). 3-D solid and finite element modeling of biomechanical structures - a software pipeline. In *EUROMECH Solid Mechanics Conference*.
- S. RIEDERER, R. BUSSE, R. GRIMM, C. JACK, D. KRUGER, P. ROSSMAN, E H. A. WARD (2001). Real-time signal processing techniques in MRI. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 3, páginas 2240–2243. IEEE.
- S. J. RIEDERER, T. TASCIYAN, F. FARZANEH, J. N. LEE, R. C. WRIGHT, E R. J. HERFKENS (1988). MR fluoroscopy: technical feasibility. *Magnetic resonance in medicine*, 8(1):1–15.
- P. ROEMER, W. EDELSTEIN, C. HAYES, S. SOUZA, E O. MUELLER (1990). The NMR phased array. *Magnetic Resonance in Medicine*, 16(2):192–225.
- N. ROFSKY, V. LEE, G. LAUB, M. POLLACK, G. KRINSKY, D. THOMASSON, E M. AMBROSINO (1999). Abdominal MR Imaging with a Volumetric Interpolated Breath-hold Examination. *Radiology*, 212(3):876–884.
- E. ROGERS (2006). A preliminary study of the acoustics of liquids in wanyi. In *11th Australasian International Conference on Speech Science e Technology, New Zealand*, páginas 427–432.
- M. ROKKAKU, K. HASHIMOTO, S. IMAIZUMI, S. NIIMI, E S. KIRITANI (1986). Measurement of the three-dimensional shape of the vocal tract based on the Magnetic Resonance Imaging technique. *Annual Bulletin of the Research Institute for Logopedics and Phoniatics, Tokyo, Japan*, 20:47–54.
- A. ROMANO E P. BADIN (2009). An MRI Study on the articulation properties of Italian consonants. *Estudios de Fonética Experimental*, XVIII:327–344.
- S. ROSSATO, G. FENG, E R. LABOISSIÈRE (1998). Recovering gestures from speech signals: a preliminary study for nasal vowels. In *Proceedings of the 5th International Conference on Spoken Language Processing, Incorporating The 7th Australian International Speech Science and Technology Conference, Sydney Convention Centre, Sydney, Australia, 30th November - 4th December 1998*. ISCA.

- S. ROSSATO, A. TEIXEIRA, E L. FERREIRA (2006). Les nasales du portugais et du français: une Étude comparative sur les donnés emma. In *XXVIes Journées d'Etudes sur la parole*.
- M. ROY, S. FOUFOU, E F. TRUCHETET (2004). Mesh comparison using attribute deviation metric. *International Journal of Image and Graphics*, 4(1):2004.
- S. RUA E D. FREITAS (2007). Morphological dynamic study of human vocal tract. In *Proceedings of the International Conference CompIMAGE-Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications, Coimbra, Portugal*.
- P. RUBIN, T. BAER, E P. MERMELSTEIN (1981). An articulatory synthesizer for perceptual research. *The Journal of the Acoustical Society of America*, 70(2):321–328.
- P. RUBIN, E. SALTZMAN, L. GOLDSTEIN, R. MCGOWAN, M. TIEDE, E C. BROWMAN (1996). CASY and extensions to the task-dynamic model. In *Proceedings of 1st ESCA Tutorial and Research Workshop on Speech Production and modeling - 4th Speech Production Seminar*, páginas 125–128.
- V. RUNGE (2007). *Clinical 3T Magnetic Resonance*. Thieme.
- B. RUTT (2003). Gradient Safety Concerns. In *Proceedings of the Intl. Soc. Mag. Reson. Med.*, página 199.
- F. RYBICKI E R. MULKERN (2001). Principles and clinical applications of single-shot Magnetic Resonance Imaging. *Supplement to Applied radiology*, 30(4):30–37.
- R. SÁ-NOGUEIRA (1938). *Elementos para um tratado de Fonética Portuguesa*. Imprensa Nacional de Lisboa, Lisboa.
- Z. SALAH, J. ORMAN, E D. BARTZ (2005). Live-wire revisited. *Bildverarbeitung für die Medizin 2005*, páginas 158–162.
- R. SAMPSON (1999). *Nasal Vowel Evolution in Romance*. Oxford University Press, New York.
- J. M. SANTOS E L. M. CHAVES (2012). A realização da lateral palatal /L/ no atlas linguístico do ACRE (ALIAC). *Revista Philologus-Suplemento: Anais da VII JNLFLP*, (54).
- M. SATO-WAKABAYASHI, M. S. INOUE-ARAI, T. ONO, E.-i. HONDA, T. KURABAYASHI, E K. MORIYAMA (2008). Combined fMRI and MRI movie in the evaluation of articulation in subjects with and without cleft lip and palate. *The Cleft Palate-Craniofacial Journal*, 45(3):309–314.
- D. J. SCHAEFER, J. D. BOURLAND, E J. A. NYENHUIS (2000). Review of Patient Safety in Time-Varying Gradient Fields. *Journal of magnetic resonance imaging*, 12(1):20–29.
- K. SCHEFFLER E S. LEHNHARDT (2003). Principles and applications of balanced SSFP techniques. *European Radiology*, 13(11):2409–2418.
- J. SCHENCK (2000). Safety of strong, static magnetic fields. *Journal of Magnetic Resonance Imaging*, 12(2):2–19.

- A. SCHENK, G. PRAUSE, E. H. PEITGEN (2000). Efficient semiautomatic segmentation of 3D objects in medical images. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000*, páginas 71–131. Springer.
- F. SCHICK (2005). Whole-body MRI at high field: technical limits and clinical potential. *European Radiology*, 15:946–959.
- I. C. SEARA (2000). *Estudo acústico-perceptual da nasalidade das vogais do português brasileiro*. Tese de Doutorado, Universidade Federal de Santa Catarina.
- N. SEIBERLICH, P. EHSES, J. DUERK, R. GILKESON, E. M. GRISWOLD (2011). Improved radial GRAPPA calibration for real-time free-breathing cardiac imaging. *Magnetic Resonance in Medicine*, 65(2):492–505.
- D. SELL E V. PEREIRA (2012). Instrumentation in the Analysis of the Structure and Function of the Velopharyngeal Mechanism. *Cleft Palate Speech: Assessment and Intervention*, página 145.
- A. SERRURIER (2006). *Modélisation tridimensionnelle des organes de la parole à partir d’images IRM pour la production de nasales: Caractérisation articulatoire-acoustique des mouvements du voile du palais*. Tese de Doutorado, Institut National Polytechnique de Grenoble.
- A. SERRURIER E P. BADIN (2005a). A Three-Dimensional Linear Articulatory Model of Velum Based on MRI Data. In *Proceedings of Interspeech, Lisboa, Portugal*.
- A. SERRURIER E P. BADIN (2005b). Towards a 3D articulatory model of the velum based on MRI and CT images. *ZAS Papers in Linguistics*, 40:195–211.
- A. SERRURIER E P. BADIN (2008). A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data. *The Journal of the Acoustical Society of America*, 123:23–35.
- C. H. SHADLE, M. A.-J. MOHAMMAD, J. N. CARTER, E. P. J. B. JACKSON (1999). Multi-planar dynamic magnetic resonance imaging: new tools for speech research. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS), San Francisco, USA*, páginas 623–626.
- R. SHAW, J. FOLKINS, E. D. KUEHN (1980). Comparisons of methods for measuring velar position from lateral view cineradiography. *Cleft Palate Journal*, 17(4):326–329.
- F. SHELLOCK (2001). *Magnetic resonance Procedures: Health Effects and Safety*. CRC Press, Florida, primeira edição.
- F. G. SHELLOCK (2000). Radiofrequency energy-induced heating during MR procedures: a review. *Journal of Magnetic Resonance Imaging*, 12:30–36.
- F. G. SHELLOCK E A. SPINAZZI (2008). MRI safety update 2008: part 2, screening patients for MRI. *American Journal of Roentgenology*, 191(4):1140–1149.

- R. SHOSTED (2011). An EMA-Aerodynamic Approach to the Velic Opening Hypothesis: Evidence from Hindi Vowel Pairs. In *Proceedings of the 17th International Congress of Phonic Sciences (ICPhS), Hong-Kong, China*.
- R. SHOSTED, C. CARIGNAN, E P. RONG (2012a). Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi. *The Journal of the Acoustical Society of America*, 131:455–465.
- R. SHOSTED, B. SUTTON, E A. BENMAMOUN (2012b). Using magnetic resonance to image the pharynx during Arabic speech: Static and dynamic aspects. In *Proceedings of Interspeech, Portland, USA*.
- A. SILVA (1995). *Análise de fonemas nasais da língua portuguesa*. Tese de Mestrado, Universidade de Aveiro, Aveiro.
- A. H. SILVA (1996). *Para a descrição fonético-acústica das líquidas no Português Brasileiro: dados de um informante Paulistano*. Tese de Mestrado, Universidade Estadual de Campinas - Unicamp, Campinas.
- J. SILVA (2005). *Segmentação Pulmonar em Estudos de Tomografia Axial Computorizada*. Tese de doutoramento, Universidade de Aveiro, Aveiro.
- J. SILVA, B. SANTOS, A. SILVA, E J. MADEIRA (2004). Modelos Deformáveis na Segmentação de Imagens Médicas: uma introdução. *Revista do DETUA*, 4(3).
- S. SILVA (2007). *Polymeco: A polygonal mesh analysis and comparison tool*. Tese de Mestrado, Universidade de Aveiro, Aveiro.
- S. SILVA (2012). *Left ventricle functional analysis from coronary CT angiography*. Tese de Doutoramento, Universidade de Aveiro, Aveiro.
- S. SILVA (2014). Integration of 3D Articulatory Model on SAPWindows. Relatório técnico, Projeto HERON II - Universidade de Aveiro, Aveiro.
- S. SILVA, J. MADEIRA, E B. SOUSA SANTOS (2009). PolyMeCo - An Integrated Environment for Polygonal Mesh Analysis and Comparison. *Computers Graphics*, 33:181–191.
- S. SILVA, P. MARTINS, C. OLIVEIRA, A. SILVA, E A. TEIXEIRA (2012). Segmentation and Analysis of the Oral and Nasal Cavities from MR Time Sequences. In A. CAMPILHO E M. KAMEL (editores), *Image Analysis and Recognition*, volume 7325 de *Lecture Notes in Computer Science*, páginas 214–221. Springer Berlin / Heidelberg. ISBN 978-3-642-31297-7.
- S. SILVA, B. S. SANTOS, C. FERREIRA., J. MADEIRA, E A. SILVA (2011). A preparatory study to choose similarity metrics for left-ventricle segmentations comparison. In *SPIE Medical Imaging 2011 – Computer-Aided Diagnosis, Orlando, USA*.
- S. SILVA E A. TEIXEIRA (2013). AAM Based Vocal Tract Segmentation from Real-Time MRI Image Sequences. In *Proceedings of the 19th Portuguese Conference on Pattern Recognition (RecPad 2013)*.

- S. SILVA E A. TEIXEIRA (2014). Systematic and Quantitative Analysis of Vocal Tract Data: Intra and Inter-Speaker Analysis. In *Proceedings of the 10th International Seminar on Speech Production (ISSP), Cologne, Germany*.
- S. SILVA, A. TEIXEIRA, C. OLIVEIRA, E P. MARTINS (2013a). Segmentation and analysis of vocal tract from midsagittal real-time MRI. In *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR), Póvoa de Varzim, Portugal*.
- S. SILVA, A. TEIXEIRA, C. OLIVEIRA, E P. MARTINS (2013b). Towards a Systematic and Quantitative Analysis of Vocal Tract Data. In *Proceedings of Interspeech, Lyon, France*.
- D. SODICKSON (2006). *Parallel Imaging Methods*, volume 1. Elsevier Inc., Philadelphia, terceira edição.
- D. SODICKSON E W. MANNING (1997). Simultaneous Acquisition of Spatial Harmonics (SMASH): Fast Imaging with Radiofrequency Coil Arrays. *Magnetic Resonance in Medicine*, 38:591–603.
- M.-J. SOLÉ (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech*, 38(1):1–23.
- M. SONKA E M. FITZPATRICK (2004). *Handbook of Medical imaging - Medical Imaging Processing and analysis*. The Society of Photo-optical Instrumentation Engineers - SPIE, segunda edição.
- M. SONKA, V. HLAVAC, E R. BOYLE (2007). *Image Processing, Analysis, and Machine Vision*. CL-Engineering. Thomson-Engineering, segunda edição.
- A. SOQUET, V. LECUIT, T. METENS, E D. DEMOLIN (2002). Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI. *Speech Communication*, 36:169–180.
- E. M. SOUZA (1994). *Para a caracterização fonético-acústica da nasalidade no Português do Brasil*. Tese de Mestrado, Instituto de Estudos Lingüísticos - Campinas.
- R. SPROAT E O. FUJIMURA (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21:291–311.
- J. STAFFORD (2004). High Field MRI: Technology, Applications, Safety, and Limitations. In *Proceedings of the AAPM 46th Annual Meeting, Pittsburgh-PA*.
- C. C. STEIN (2011). O percurso acústico-articulatório da alofonia da consoante lateral palatal. *Domínios de Linguagem*, 5(1):219–246.
- K. STEVENS (1998). *Acoustic Phonetics*. The MIT Press, Cambridge, Massachusetts EUA.
- K. STEVENS, A. ANDRADE, E M. VIANA (1987). Perception of vowel nasalization in VC contexts: A cross-language study. *Journal of the Acoustical Society of America*, 82(S1, S119).
- K. STEVENS E S. OHMAN (1963). Cineradiographic studies of speech. *STL-QPSR*, 4(2):009–011.

- M. STONE (1999). *Laboratory Techniques for Investigating Speech Articulation*, páginas 11–32. Blackwell.
- M. STONE (2005). A Guide to Analyzing Tongue Motion from Ultrasound Images. *Clinical linguistics and phonetics*, 19(6-7):455–502.
- M. STONE E A. J. LUNDBERG (1996). Three-dimensional tongue surface shapes of english consonants and vowels. *Journal of the Acoustical Society of America (JASA)*, 99(6):3728–3737.
- B. H. STORY, I. R. TITZE, E E. A. HOFFMAN (1996). Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America (JASA)*, 100(1):537–554.
- P. STREVEVS (1954). Some observations on the phonetics and pronunciation of mode Portuguese. *Revista do Laboratório de Fonética Experimental (Coimbra)*, II:5–29.
- J. SUNDBERG, C. JOHANSSON, H. WILBRAND, E C. YTTERBERGH (1987). From sagittal distance to area. *Phonetica*, 44(2):76–90.
- Y. SUO, E. Z. MURANO, M. STONE, P. M. EVITTS, E J. L. PRINCE (2011). Semi-automatic Segmentation of Speech Articulators Using MRI Images. In *9th International Seminar on Speech Production (ISSP), Montreal, Canadá*, páginas 9–16.
- S. TAKANO E K. HONDA (2007). An MRI analysis of the extrinsic tongue muscles during vowel production. *Speech Communication*, 49(1):49–58.
- H. TAKEMOTO (2001). Morphological analyses of the human tongue musculature for three-dimensional modeling. *Journal of Speech, Language, and Hearing Research*, 44:95–107.
- H. TAKEMOTO, K. HONDA, S. MASAKI, Y. SHIMADA, E I. FUJIMOTO (2003). Measurement of Temporal Changes in Vocal Tract Area Function during a continuous vowel sequence using a 3D Cine-MRI Technique. In *Proceedings of the 6th International Seminar on Speech Production, Sydney, Austrália*, páginas 284–289.
- H. TAKEMOTO, K. HONDA, S. MASAKI, Y. SHIMADA, E I. FUJIMOTO (2006). Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. *Journal of Acoustical Society of America*, 119(2):1037–1049.
- H. TAKEMOTO, T. KITAMURA, H. NISHIMOTO, E K. HONDA (2004). A method of tooth superimposition of MRI data for accurate measurement of vocal tract shape and dimensions. *Acoustical Science and Technology*, 25(6):468–474.
- P. TAYLOR (2009). *Text-to-speech synthesis*, volume 15. Cambridge University Press.
- A. TEIXEIRA (2000). *Síntese Articulatoria das Vogais Nasais do Português Europeu*. Tese de doutoramento, Universidade de Aveiro, Aveiro.
- A. TEIXEIRA, R. MARTINEZ, L. N. SILVA, L. M. T. JESUS, J. C. PRÍNCIPE, E F. A. C. VAZ (2005). Simulation of Human Speech Production Applied to the Study and Synthesis of European Portuguese. *EURASIP Journal on Applied Signal Processing*, 9:1435–1448.

- A. TEIXEIRA, P. MARTINS, C. OLIVEIRA, C. FERREIRA, A. SILVA, E R. SHOSTED (2012a). Real-Time MRI for Portuguese. In H. CASELI, A. VILLAVICENCIO, A. TEIXEIRA, E F. PERDIGÃO (editores), *Computational Processing of the Portuguese Language*, volume 7243 de *Lecture Notes in Computer Science*, páginas 306–317. Springer Berlin / Heidelberg. ISBN 978-3-642-28884-5.
- A. TEIXEIRA, P. MARTINS, C. OLIVEIRA, E A. SILVA (2011). An MRI study of consonantal coarticulation resistance in Portuguese. In *Proceedings of the 9th International Seminar on Speech Production (ISSP), Montreal, Canadá*, páginas 243–250.
- A. TEIXEIRA, P. MARTINS, C. OLIVEIRA, E A. SILVA (2012b). Production and Modeling of the European Portuguese Palatal Lateral. In H. CASELI, A. VILLAVICENCIO, A. TEIXEIRA, E F. PERDIGÃO (editores), *Computational Processing of the Portuguese Language*, volume 7243 de *Lecture Notes in Computer Science*, páginas 318–328. Springer Berlin / Heidelberg. ISBN 978-3-642-28884-5.
- A. TEIXEIRA, C. OLIVEIRA, P. MARTINS, I. DOMINGUES, E A. SILVA (2009). The project HERON. In *Proceedings of the I SLTech in Iberian languages, Oeiras, Portugal*.
- A. TEIXEIRA E F. VAZ (2001). European Portuguese Nasal Vowels: An EMMA Study. In *Proceedings of the European Conference on Speech Communication and Technology - Eurospeech*.
- A. TEIXEIRA, F. VAZ, E J. C. PRÍNCIPE (2000). Nasal Vowels Following a Nasal Consonant. In *Proceedings of the 5th Seminar on Speech Production: Models and Data, Kloster Seeon, Bavaria, Germany*, páginas 285–288.
- J. P. TEIXEIRA, D. FREITAS, D. BRAGA, M. J. BARROS, E V. LATSCH (2001). Phonetic events from the labeling the european portuguese database for speech synthesis, FEUP/IPBDB. In *Proceedings of Interspeech2001, Aalborg, Denmark*, páginas 1707–1710.
- J. M. THEYSOHN, S. MADERWALD, O. KRAFF, C. MOENNINGHOFF, M. E. LADD, E S. C. LADD (2008). Subjective acceptance of 7 Tesla MRI for human imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 21(1-2):63–72.
- W. TIAN, H. YIN, R. J. REDETT, B. SHI, J. SHI, R. ZHANG, E Q. ZHENG (2010). Magnetic resonance imaging assessment of the velopharyngeal mechanism at rest and during speech in Chinese adults and children. *Journal of Speech, Language and Hearing Research*, 53(6):1595.
- M. TIEDE, S. MASAKI, E E. VATIKIOTIS-BATESON (2000). Contrasts in speech articulation observed in sitting and supine conditions. In *Proceedings of the 5th Seminar on Speech Production, Kloster Seeon, Bavaria, Germany*, páginas 25–28.
- K. TOM, I. TITZE, E. HOFFMAN, E B. STORY (1999). 3-D Vocal Tract Imaging and Formant Structure: Varying Vocal Register, Pitch and Loudness. *NCVS Status and Progress Report*, (14):101–113.
- J. TSAO, P. BOESIGER, E K. PRUESSMANN (2003). k-t BLAST and k-t SENSE: Dynamic MRI with High Frame Rate Exploiting Spatiotemporal Correlations. *Magnetic Resonance in Medicine*, 50:1031–1042.

- J. TSAO E S. KOZERKE (2012). MRI temporal acceleration techniques. *Journal of Magnetic Resonance Imaging*, 36(3):543–560.
URL <http://dx.doi.org/10.1002/jmri.23640>
- O. TYMOFIYEVA, K. ROTTNER, D. GAREIS, J. BOLDT, F. SCHMID, M. LOPEZ, E.-J. RICHTER, E P. JAKOB (2008). In vivo MRI-based dental impression using an intraoral RF receiver coil. *Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering*, 33(4):244–251.
- M. UECKER, S. ZHANG, E J. FRAHM (2010a). Nonlinear inverse reconstruction for real-time MRI of the human heart using undersampled radial FLASH. *Magnetic Resonance in Medicine*, 63(6):1456–1462.
URL <http://dx.doi.org/10.1002/mrm.22453>
- M. UECKER, S. ZHANG, D. VOIT, A. KARAS, K.-D. MERBOLDT, E J. FRAHM (2010b). Real-time mri at a resolution of 20 ms. *NMR in Biomedicine*, 23(8):986–994.
URL <http://dx.doi.org/10.1002/nbm.1585>
- K. VAN DEN DOEL, F. VOGT, R. ENGLISH, E S. FELS (2006). Towards articulatory speech synthesis with a dynamic 3D finite element tongue model. In *Proceedings of the 7th International Seminar on Speech Production, Ubatuba, Brasil*.
- M. VASCONCELOS, S. VENTURA, D. FREITAS, E J. M. TAVARES (2010). Using statistical deformable models to reconstruct vocal tract shape from magnetic resonance images. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 224(4).
- S. VENTURA, D. FREITAS, E J. M. TAVARES (2008). Three-dimensional modeling of the tongue during speech using MRI data. In *Proceedings of 8th International Symposium on Computer Methods in Biomechanics and Biomedical Engineering, Porto, Portugal*.
- S. M. VENTURA, D. R. S. FREITAS, I. M. A. RAMOS, E J. M. R. TAVARES (2013). Morphologic Differences in the Vocal Tract Resonance Cavities of Voice Professionals: An MRI-Based Study. *Journal of Voice*, páginas 132–140.
- A. VIANA (1883). *Essai de phonétique et phonologie de la langue portugaise d'après le dialecte actuel de Lisbonne*. Lisboa.
- F. VOGT, C. WILSON, R. CHANDER, E S. FELS (2009). Extraction of tongue shapes from magnetic resonance Images. In *OPAL, Vancouver, Canadá*.
- H. WANG, D. LIANG, K. F. KING, G. NAGARSEKAR, Y. CHANG, E L. YING (2012). Improving grappa using cross-sampled autocalibration data. *Magnetic Resonance in Medicine*, 67(4):1042–1053.
URL <http://dx.doi.org/10.1002/mrm.23083>
- J. C. WELLS (1982). *Accents of English*, volume 1. Cambridge University Press.
- J. C. WELLS (1995). Computer-coding the IPA: a proposed extension of SAMPA. *Revised draft*, 4(28):1995.

- H. F. WERTZNER, L. d. O. PAGAN-NEVES, E M. M. d. CASTRO (2007). Acoustic analysis and stimulability index for Brazilian Portuguese liquid sounds. *Revista CEFAC*, 9(3):339–350.
- O. WIEBEN (2007). *Parallel imaging in clinical MR applications*, capítulo Complementary Techniques for Accelerated Imaging, páginas 91–103. Springer Berlin/Heidelberg.
- O. WIRJADI (2007). *Survey of 3d image segmentation methods*. Fraunhofer ITWM.
- A. WISMUELLER, J. BEHREND, P. HOOLE, G. LEINSINGER, M. REISER, E P.-L. WESTESSON (2008). Human Vocal Tract Analysis by in Vivo 3D MRI during Phonation: A Complete System for Imaging, Quantitative Modeling, and Speech Synthesis. In D. METAXAS, L. AXEL, G. FICHTINGER, E G. SZÉKELY (editores), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, volume 5242 de *Lecture Notes in Computer Science*, páginas 306–312. Springer Berlin / Heidelberg. ISBN 978-3-540-85989-5.
- D. WITHEY E Z. KOLES (2008). A Review of Medical Image Segmentation: Methods and Available Software. *International Journal of Bioelectromagnetism*, 10(3):125–148.
- A. A. WRENCH E J. M. SCOBIE (2003). Categorising Vocalisation of English /l/ using EPG, EMA and Ultrasound. In *Proceedings of the 6th International Seminar on Speech Production (ISSP)*, Sydney Australia, páginas 314–319.
- A. A. WRENCH E J. M. SCOBIE (2011). Very high frame rate ultrasound tongue imaging. In *Proceedings of the 9th International Seminar on Speech Production (ISSP)*, Montreal, Canadá, páginas 155–162.
- C. XU, D. PHAM, J. PRINCE, *et al.* (2000). Image segmentation using deformable models. *Handbook of medical imaging*, 2:129–174.
- B. YANG (1999). Measurement and synthesis of the vocal tract of Korean monophthongs by MRI. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco, USA, páginas 2005–2008.
- P. A. YUSHKEVICH, J. PIVEN, H. C. HAZLETT, R. G. SMITH, S. HO, J. C. GEE, E G. GERIG (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128.
- N. ZARKOVA, N. HEWLETT, E W. HARDCASTLE (2008). An Ultrasound study of lingual coarticulation in children and adults. In *Proceedings of the 8th International Seminar on Speech Production (ISSP)*, Strasbourg, France.
- J. P. ZERLING (1984). Phenomenes de nasalité et de nasalization vocaliques: Étude cineradiographique pour deux locuteurs. *Travaux de l'Institut de Phonétique de Strasbourg*, 16:241–266.
- S. ZHANG, K. T. BLOCK, E J. FRAHM (2010a). Magnetic resonance imaging in real time: Advances using radial FLASH. *Journal of Magnetic Resonance Imaging*, 31(1):101–109.

- S. ZHANG, A. OLTHOFF, E J. FRAHM (2012). Real-time magnetic resonance imaging of normal swallowing. *Journal of Magnetic Resonance Imaging*, 35(6):1372–1379.
URL <http://dx.doi.org/10.1002/jmri.23591>
- S. ZHANG, M. UECKER, D. VOIT, K.-D. MERBOLDT, E J. FRAHM (2010b). Real-time cardiovascular magnetic resonance at high temporal resolution: radial FLASH with nonlinear inverse reconstruction. *Journal of Cardiovascular Magnetic Resonance*, 12(1):39.
URL <http://www.jcmr-online.com/content/12/1/39>
- Z. ZHANG E C. ESPY-WILSON (2004). A vocal tract model of American English /l/. *Journal of the Acoustical Society of America*, 115:1274–1280.
- Z. ZHANG, C. ESPY-WILSON, E M. TIEDE (2003). Acoustic modeling of of American English Lateral Approximants. In *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech)*, Geneva, Switzerland, páginas 2393–2396.
- X. ZHOU (2009). *An MRI-Based articulatory and acoustic study of American English liquid souns /r/ and /l/*. Tese de doutoramento, University of Maryland, USA.
- X. ZHOU, C. ESPY-WILSON, M. TIEDE, E S. BOYCE (2010). An MRI based articulatory and acoustic study of lateral sounds in American English. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, USA.
- X. ZHOU, J. WOO, M. STONE, E C. ESPY-WILSON (2013). A Cine MRI-Based Study Of sibilant Fricatives Production in Post-Glossectomy Speakers. In *Proceedings of Interspeech*, Lyon, France.
- S. C. ZHU E A. YUILLE (1996). Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900.
- Y. ZHU, Y.-C. KIM, M. I. PROCTOR, S. S. NARAYANAN, K. S. NAYAK, *et al.* (2013a). Dynamic 3-D visualization of vocal tract shaping during speech. *IEEE Transactions on Medical Imaging*, 32(5):838–848.
- Y. ZHU, A. TOUTIOS, S. NARAYANAN, E K. NAYAK (2013b). Faster 3D Vocal Tract Real-time MRI Using Constrained Reconstruction. In *Proceedings of Interspeech*, Lyon, France.
- J. ZHUO E R. P. GULLAPALLI (2006). MR Artifacts, Safety, and Quality Control. *Radiographics*, 26(1):275–297.
- A. ZIERDT, P. HOOLE, E H. G. TILLMANN (1999). Development of a System for Three-Dimensional Fleshpoint Measurement of Speech Movements. In J. J. OHALA, Y. HASEGAWA, M. OHALA, D. GRANVILLE, E A. C. BAILEY (editores), *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco, USA, páginas 73–75.

- K. ZOU, S. WARFIELD, A. BHARATHA, C. TEMPANY, M. KAUS, S. HAKER, W. WELLS, F. JOLESZ, E R. KIKINIS (2004). Statistical validation of image segmentation quality based on a spatial overlap index: scientific reports. *Academic radiology*, 11(2):178–189.

Lista de Siglas e Acrónimos

AAM	<i>Active Appearance Models</i>
ANOVA	<i>Analysis of Variance</i>
ASM	<i>Active Shape Models</i>
BW	<i>Bandwidth</i>
CASY	<i>Configurable Articulatory Synthesizer</i>
CAIPIRINHA	<i>Controlled Aliasing In Parallel Imaging Results IN Higher Acceleration</i>
C/R	Relação Contraste-Ruído - no inglês <i>Contrast to Noise Ratio</i>
dB/dt	Campo magnético variável
DICOM	<i>Digital Imaging and Communications in Medicine</i>
DFT	<i>Discret Fourier Transform</i>
DWI	<i>Diffusion Weighted Imaging</i>
EBCT	<i>Electron Beam Computed Tomography</i>
ECG	Electrocardiograma
EG	Eco de Gradiente
EMA	<i>ElectroMagnetic Articulography</i> ou Articulografia Electromagnética.
ENP	Estimulação Nervosa Periférica
EPG	Electropalatografia
EPI	<i>Echo-Planar Imaging</i> ou Imagem Eco Planar
ERM	Espetroscopia por Ressonância Magnética
ES	Eco de Spin
ESP	<i>Echo Spacing</i>
ETL	<i>Echo Train Length</i> ou turbo factor
EURATOM	<i>European Atomic Energy Community</i>
FA	<i>Flip Angle</i>

- FDA** *Food and Drug Administration*
- FID** *Free Induction Decay*
- FIESTA** *Fast Imaging Employing Steady State Acquisition*
- FISP** *Fast Imaging with Steady Precession*
- FLASH** *Fast Low Angle Shot*
- FOV** *Field of View* ou campo de visão
- FSE** *Fast Spin Echo*
- FFE** *Fast Field Echo*
- FFT** *Fast Fourier Transform*
- GUI** *Graphical User Interface*
- GRAPPA** *Generalized Autocalibrating Partially Parallel Acquisition*
- GRASS** *Gradient Recalled Acquisition in the Steady State*
- HATS** *Head and Transducer Support System*
- HASTE** *Half-Fourier Single-Shot Turbo Spin Echo*
- HF** *Half-Fourier*
- HPA** *Health Protection Agency*
- IBILI** Instituto Biomédico de Investigação da Luz e Imagem
- ICNIRP** *International Commission on Non-Ionizing Radiation Protection*
- ICP** *Institute de la Communication Parléé*
- IEC** *International Electrotechnical Commission*
- IEETA** Instituto de Eletrónica e Telemática da Universidade de Aveiro
- IP** Índice de Pratt
- IS** Intensidade de Sinal
- ITK** *Interactive ToolKit*
- KTH** *Kungl Tekniska Högskolan* - Instituto Real de Tecnologia de Estocolmo.
- LAVA** *Liver Acquisition in Volumetric Acquisition*
- Lolo** *Local Lock*
- MIT** *Massachusetts Institute of Technology*
- MHRA** *Medicine and Healthcare products Regulatory Agency*
- MO** Microfone Ótico
- MP-RAGE** *Magnetization Prepared-Rapid Acquired Gradient Echo*
- MRI** *Magnetic Resonance Imaging*
- NEX** *Number of Excitations*

OPG Optopalatografia

PACS *Picture Archiving and Communication System*

PB Português Brasileiro ou Português do Brasil

PCA *Principal Component Analysis*

PE Português Europeu

PI *Parallel Imaging* ou técnicas de imagem paralela

PILS *Partially parallel Imaging with Localized Sensitivities*

PVF Passagem Velofaríngea ou Porto Velofaríngeo

R Fator de aceleração

RARE *Rapid Acquisition with Refocused Echoes ou Rapid Acquisition with Relaxation Enhancement*

RF Radiofrequência

RG *Region Growing*

RM Ressonância Magnética

RMf Ressonância Magnética funcional

ROI *Region of Interest* ou Região de Interesse

SAMPA *Speech Assessment Methods Phonetic Alphabet*

SAPWindows Sintetizador Articulatorio do Português para plataforma Windows

SAR *Specific Absorption Rate*

SE *Spin Echo* ou Eco de Spin

SENSE *Sensitivity Encoding*

SMASH *Simultaneous Acquisition of Spacial Harmonics*

SPACE-RIP *Sensitivity Profiles from an Array of Coils for Encoding and Reconstruction In Parallel*

SPAN *Speech Production and Articulation kNowledge Group*

S/R Relação Sinal-Ruído - do inglês *Signal to Noise Ratio*(SNR)

SS *Single-Shot*

SSM *Synchronized Sampling Method*

SSFP *Steady State Free Precession*

SS-TSE *Single-Shot Turbo Spin Echo ou Fast Spin Echo*

SWI *Susceptibility Weighted Imaging*

TE Tempo de Eco

TC Tomografia Computorizada

THRIVE *T1 High Resolution Isotropic Volume Excitation*

TR Tempo de Repetição

TF Transformada de *Fourier*

TrueFISP *True Fast Imaging with Steady Precession*

TSE *Turbo Spin Echo*

TTL *Transistor-Transistor Logic*

TTLag *Target-to-Target Lag*

T2TLag *Target2-to-Target Lag*

TTS *Text-to-Speech*

UA Universidade de Aveiro

UCLA *University of California, Los Angeles*

UPSID *UCLA Phonological Segment Inventory Database*

US Ultrasonografia

USC *University of Southern California*

VIBE *Volume Interpolated Breath-hold Examination*

VOI *Volume of Interest*

VPO *Velum Port Opening*

VTAR *Vocal Tract Acoustic Response*

VTK *Visualization Toolkit*

XRMB *X-Ray Microbeam*

X-SAMPA *Extended Speech Assessment Methods Phonetic Alphabet*

Apêndice A

Informação ao Participante

Documento elaborado seguindo as orientações gerais constantes do documento:
CIOMS (2002). *International ethical guidelines for biomedical research involving human subjects*.
Council for International Organizations of Medical Sciences - World Health Organization.



INFORMAÇÃO AO PARTICIPANTE

Título do Trabalho de Investigação – Ressonância Magnética no Estudo da Produção do Português Europeu*

Instituição – Universidade de Aveiro /Instituto de Electrónica e Telemática da UA (IEETA)

* Este estudo está incluído no Projecto aprovado pela FCT – Ref: PTDC/EEA-PLP/098298/2008- HERON II Evolução do Sistema de Síntese de Voz de Base articulatória para o Português Europeu.

Investigador – Paula Maria Vaz Martins

Responsáveis Científicos – Professor Doutor António Joaquim da Silva Teixeira* / Professor Doutor Augusto Marques Ferreira da Silva

* Investigador Principal do Projecto HERON II

Objectivo da investigação

O objectivo principal desta investigação, enquadrada num projecto mais vasto, O PROJECTO HERON II, visa a recolha de dados anatómicos e articulatórios, com protocolos RM ajustados à produção de fala, com o intuito de aprofundar o conhecimento a nível da produção do Português Europeu e desenvolver e aperfeiçoar um sintetizador de voz de base articulatória existente na Universidade de Aveiro (SAP-Windows).

Metodologias

A sua participação neste estudo envolverá a aquisição de imagens com Ressonância Magnética, durante a qual serão efectuadas algumas sequências com o objectivo de obter imagens representativas da configuração do tracto vocal, durante a produção dos vários sons do PE. Poderá ser também gravado o sinal de voz durante ou após a aquisição de imagens. O tempo total de aquisição (RM) será de cerca de 1 hora. Durante as aquisições vai-lhe ser solicitado que sustenha ou produza de forma natural um determinado som do PE. Antes das aquisições de RM efectuará um teste com um Terapeuta da Fala qualificado para avaliação do tipo de oclusão dentária e para confirmação da inexistência de patologias articulatórias.

Características do estudo RM

A Ressonância Magnética (RM) é uma modalidade de imagem, muito utilizada na prática clínica, sendo considerado um exame seguro, desde que observadas algumas normas segurança. Por essa razão, será devidamente informado das contra-indicações e dos procedimentos que deverá observar antes da entrada na sala de RM. Ser-lhe-á efectuado, por um técnico qualificado, um inquérito de rastreio. Este inquérito tem como objectivo assegurar que não tem nenhum material no interior do seu corpo que seja contra-indicado para a realização de RM. Nessa situação a sua participação no estudo seria cancelada.

O processo de aquisição de imagem por RM, envolve a utilização de um campo magnético estático de elevada intensidade, e de campos de Radiofrequência (RF).

O equipamento utilizado para o estudo apresenta um magneto em formato de túnel, aberto nas duas extremidades e bem ventilado. Mesmo que exista alguma falha de corrente é possível retirá-lo com facilidade. Para realizar o exame será colocado em posição de decúbito dorsal, confortável e com a cabeça colocada num dispositivo chamado antena ou bobine. A sua função é captar e emitir pulsos de RF.

Durante a aquisição é absolutamente fundamental que não movimente a cabeça. Esta será imobilizada para que fique numa posição confortável. Ser-lhe-ão fornecidos auscultadores para minimizar a intensidade do ruído (normal nas aquisições de RM) e para facilitar a comunicação durante o procedimento. Terá na sua mão uma campainha que deverá apertar caso se sinta desconfortável ou necessite de comunicar com a equipa.

Confidencialidade

Será sempre salvaguardada a confidencialidade dos dados adquiridos, nunca aparecendo o seu nome nas imagens. Nas imagens serão utilizadas siglas que não permitirão a sua identificação. No entanto, será necessário fornecer alguma informação como idade, peso, altura, nacionalidade, naturalidade e habilitações académicas.

Os dados obtidos poderão ser utilizados no desenvolvimento de Dissertações de Mestrado ou Doutoramento e em publicações em revistas da especialidade.

Organização e financiamento do estudo

Este estudo é financiado pela Fundação para a Ciência e Tecnologia (que é a principal entidade financiadora de investigação em Portugal) ao abrigo do Projecto FCT (PTDC/EEA-PLP/098298/2008), decorre na Universidade de Aveiro em colaboração com entidades públicas

(IBILI) ou privadas (IMACENTRO), devidamente acreditadas, que prestem serviços de Ressonância Magnética.

Contactos para mais informações sobre o estudo

Para mais informações acerca do estudo poderá contactar:

Paula Martins
Escola Superior de Saúde da Universidade de Aveiro,
Universidade de Aveiro,
Campus Universitário de Santiago,
Edifício III, 3810-193, Aveiro
Telefone: 234 372464 /966633458

Muito obrigada !!

Apêndice	B
----------	----------

Registo para Caracterização do
Participante



universidade de aveiro

CARACTERIZAÇÃO DO FALANTE

NOME _____

IDADE _____

SIGLA _____

NATURALIDADE _____

RESIDÊNCIA _____

HABILITAÇÕES LITERÁRIAS _____

PROFISSÃO _____

PESO _____ Kg

ALTURA _____ cm

BILINGUE

Sim ☐Não ☐

CONHECIMENTOS DE FONÉTICA

Sim ☐Não ☐

TREINO VOCAL

Sim ☐Não ☐

ALTERAÇÕES DA ARTICULAÇÃO

Sim ☐Não ☐

CLASSE OCLUSAL

Tipo _____

NOTAS _____

Apêndice **C**

Declaração de Consentimento Informado



universidade de aveiro

CONSENTIMENTO INFORMADO

Título do trabalho de investigação - Ressonância Magnética no Estudo da Produção do Português Europeu*

Instituição – Universidade de Aveiro /Instituto de Electrónica e Telemática da UA (IEETA)

*Este estudo está incluído no Projecto aprovado pela FCT – Ref: PTDC/EEA-PLP/098298/2008- HERON II Evolução do Sistema de Síntese de Voz de Base Articulatoria para o Português Europeu.

Investigador - Paula Maria Vaz Martins

Responsáveis Científicos – Professor Doutor António Joaquim da Silva Teixeira* e Professor Doutor Augusto Marques Ferreira da Silva

* Investigador Principal do Projecto HERON II

ATENÇÃO

A sua participação é voluntária podendo escolher não participar no estudo ou revogar este consentimento em qualquer altura da sessão de aquisição de dados.

Qualquer dúvida que tenha deve ser colocada à equipa de investigadores.

Consentimento de participação

Eu _____ (Nome Completo)

- 1) Declaro que li e percebi o conteúdo da folha de Informação que me foi cedida pelos investigadores desta Investigação/projecto;
- 2) Confirmo que me foi efectuado um questionário de rastreio antes das aquisições de RM;
- 3) Declaro que tive hipótese de esclarecer todas as questões que me surgiram relativas ao procedimento;
- 4) Autorizo a divulgação dos dados obtidos, no âmbito de trabalhos científicos, desde que salvaguardada a confidencialidade dos mesmos.

Desta forma, dou o meu consentimento para participação no estudo acima referido.

Assinatura _____

Data ____/____/____

Investigador _____

Data ____/____/____

Investigador Principal _____

Data ____/____/____

Apêndice **D**

Questionário Rastreo de Ressonância Magnética

O questionário apresentado é o utilizado, por rotina, na unidade de RM onde foi efetuada a aquisição de dados - IBILI. Respeita as diretivas de organismos internacionais (ICNIRP, 2004; FDA, 2003) e literatura especializada no que respeita à segurança no setor de RM (Kanal *et al.*, 2007; Sherlock, 2001; Sherlock e Spinazzi, 2008).

Formulário de Segurança de Imagem por Ressonância Magnética

Projecto nº:
Exame nº:
(preencher pelo serviço)

Apelido: _____ Nome: _____ Altura: _____ cm Peso: _____ kg
Data nascimento: ____/____/____ Telefone: _____ E-mail: _____
Endereço: _____ Localidade: _____
Código Postal: _____ Concelho: _____
Nome de Contacto de urgência: _____ Telefone: _____
Médico: _____ Endereço: _____ Telefone: _____

1. Alguma vez foi submetida a alguma cirurgia e/ou algum procedimento invasivo? ☐ Sim ☐ Não (se sim, especifique em baixo)
Tipo: _____ Data: ____/____/____
Tipo: _____ Data: ____/____/____
2. Alguma vez realizou algum exame de Imagem por Ressonância Magnética? ☐ Sim ☐ Não (se sim, especifique em baixo)
Área do corpo: _____ Data: ____/____/____ Local: _____
Área do corpo: _____ Data: ____/____/____ Local: _____
3. Trabalhou como maquinista, com metal, ou lida com metal habitualmente? ☐ Sim ☐ Não
Teve algum ferimento no olho com metal (ex. lascas metálicas ou corpo estranho)? ☐ Sim ☐ Não
4. Está grávida, poderá estar grávida, ou encontra-se a amamentar? ☐ Sim ☐ Não
5. Sofre de anemia falciforme ou talassemia? ☐ Sim ☐ Não
6. Tem pacemaker ou desfibrilhador cardíaco implantado? ☐ Sim ☐ Não

A Imagem por Ressonância Magnética (IRM) utiliza um campo magnético muito elevado, gradientes de campo magnético rapidamente modificados e utilização de elevadas radiofrequências. Alguns objectos metálicos e electromagnéticos podem interferir com a realização do exame e até serem perigosos. Antes de ser autorizado a entrar, temos de saber se possui algum objecto metálico no seu corpo, aparelho electromagnético ou se se enquadra em algumas das circunstâncias abaixo referidas.

Por favor, responda de forma correcta.

- | | |
|---|--|
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Clip de aneurisma ou clip cerebral | <input type="checkbox"/> Sim <input type="checkbox"/> Não Fragmentos metálicos (ex. olho, crânio, corpo) |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Clamp vascular na artéria carótida | <input type="checkbox"/> Sim <input type="checkbox"/> Não Clip aórtico |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Neuro-estimulador | <input type="checkbox"/> Sim <input type="checkbox"/> Não Implantes metálicos ou em malha de arame |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Bomba de insulina ou de infusão | <input type="checkbox"/> Sim <input type="checkbox"/> Não Agraços cirúrgicos ou suturas |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Implante coclear, canais auditivos ou implante auditivo | <input type="checkbox"/> Sim <input type="checkbox"/> Não Barras de Harrington (coluna) |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Próteses (olho/órbita, etc) | <input type="checkbox"/> Sim <input type="checkbox"/> Não Fixador, parafuso ou placa no osso/articulação |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Implante colocado por ímã forte | <input type="checkbox"/> Sim <input type="checkbox"/> Não Peruca (remover antes de entrar) |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Próteses de válvulas cardíacas | <input type="checkbox"/> Sim <input type="checkbox"/> Não Cabelo postição |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Membro ou articulação artificial | <input type="checkbox"/> Sim <input type="checkbox"/> Não Aparelho auditivo (remover antes de entrar) |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Outros implantes no corpo ou na cabeça | <input type="checkbox"/> Sim <input type="checkbox"/> Não Dentaduras (remover antes de entrar) |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Eléctrodos (corpo, cabeça ou cérebro) | <input type="checkbox"/> Sim <input type="checkbox"/> Não Implantes dentários |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Stents intravasculares, filtros ou outros dispositivos semelhantes | <input type="checkbox"/> Sim <input type="checkbox"/> Não Asma ou doenças respiratórias |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Shunt (Intraventricular ou cerebral) | <input type="checkbox"/> Sim <input type="checkbox"/> Não Tonturas, epilepsia ou descoordenação motora |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Cateteres ou porta de acesso vascular | <input type="checkbox"/> Sim <input type="checkbox"/> Não Hospitalização por problemas mentais ou neurológicos |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Dispositivo intra-uterino ou diafragma | <input type="checkbox"/> Sim <input type="checkbox"/> Não Traumatismo craniano |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Adesivos ou pensos terapêuticos (ex. Nicotina, Controlo Gravidez, dor, etc) | <input type="checkbox"/> Sim <input type="checkbox"/> Não Enxaqueca ou dor de cabeça migratória |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Estilhaços metálicos ou balas | <input type="checkbox"/> Sim <input type="checkbox"/> Não Ataques de pânico |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Tatuagens | <input type="checkbox"/> Sim <input type="checkbox"/> Não Enfarte ou acidente vascular cerebral |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Sombra ocular (remover antes de entrar) | <input type="checkbox"/> Sim <input type="checkbox"/> Não Problemas de saúde ao deitar-se de costas |
| <input type="checkbox"/> Sim <input type="checkbox"/> Não Piercings (remover antes de entrar) | <input type="checkbox"/> Sim <input type="checkbox"/> Não Problema em terminar exame de IRM prévio |
| | <input type="checkbox"/> Sim <input type="checkbox"/> Não Claustrofobia |

Por favor remova todos os objectos metálicos antes de entrar na sala de IRM incluindo: chaves, alfinetes para o cabelo, brincos, relógios, colares, pulseiras, canetas, cintos, botões metálicos, adereços metálicos (ex. broches, pins, etc), clips, moedas, faca de bolso e roupa com metal na sua constituição. É necessária protecção auditiva durante o exame de IRM.

Confirmando que a informação acima está correcta, de acordo com o meu melhor conhecimento. Li e compreendi todas as questões e termos referidos neste formulário. Foi-me concedida a oportunidade de colocar todas as questões que achei necessárias e as minhas dúvidas acerca deste formulário foram todas esclarecidas.

Assinatura: _____ Data: ____/____/____
Conferido por: _____ Data: ____/____/____

Apêndice E

Artigo - IJCICG

P. Martins, S. Oliveira, C. and Silva, C. Ferreira, A. Silva, e A. Teixeira (2012). Polygonal Mesh Comparison applied to the study of European Portuguese Sounds. *International Journal of Creative Interfaces and Computer Graphics-IJCICG*, 3:28–44.

28 International Journal of Creative Interfaces and Computer Graphics, 3(1), 28-44, January-June 2012

Polygonal Mesh Comparison Applied to the Study of European Portuguese Sounds

Paula Martins, University of Aveiro, Portugal

Samuel Silva, University of Aveiro, Portugal

Catarina Oliveira, University of Aveiro, Portugal

Carlos Ferreira, Portuguese National Functional Brain Imaging Research Network, Portugal

Augusto Silva, University of Aveiro, Portugal

António Teixeira, University of Aveiro, Portugal

ABSTRACT

The purpose of the authors' study was to evaluate the feasibility of using a mesh comparison tool in the study of European Portuguese speech sounds. A large 3D MRI database from several speakers, including various sounds and contexts has been acquired. Segmentation, visualization, and analysis of such a large database are complex, time-consuming tasks, preventing the use of manual segmentation techniques. A more efficient semi-automatic method was devised to accomplish that task. After tongue segmentation, meshes were created from the segmented volumes and polygonal mesh comparison was used to assess differences between different sounds, vocalic contexts, syllabic positions and speakers. This is the first study using such an approach to analyze and compare tongue shape. This comparison method provides a qualitative measure enabling further insight into the main differences between speech sounds.

Keywords: 3D MRI, European Portuguese, Polygonal Mesh Comparison, PolyMeCo, Segmentation, Speech Production

INTRODUCTION

Functional and morphological modeling of the vocal tract structures is of special relevance in the fields of speech production, speech synthesis

and speech disorders. Biomechanical models of the oral, laryngeal and pharyngeal structures have been developed, in recent years, with a relevant contribution, not only for a better understanding of speech production mechanisms (e.g., Gérard, Perrier, & Payan, 2006; Perrier, Payan, Zandipour, & Perkell, 2003), but also

DOI: 10.4018/jcicg.2012010103

for the diagnoses and treatment of speech and sleep disorders (e.g., Fels et al., 2009).

In the speech synthesis field, articulatory synthesis (e.g., Birkholz & Kröger, 2006) is one of the most promising techniques (Shadle & Damper, 2001) fulfilling the requisites for an “ideal” speech synthesizer (e.g., to easily produce different voice qualities). Articulatory synthesis produces speech using models of physical, anatomical and physiological characteristics of the human production system. This technique models the human vocal tract directly, instead of modeling the signal or its acoustic characteristics (Teixeira, Oliveira, & Barbosa, 2008). As a consequence, this type of anthropomorphic synthesizers requires large amounts of detailed anatomic-physiological information, if possible in 3D (Teixeira et al., 2005).

From a phonetic/linguistic point of view, articulatory information and 3D models are essential for improving knowledge concerning the articulation of the different sounds. It can also be used as an auxiliary tool when teaching, e.g., foreign languages.

Instrumental Techniques

Obtaining accurate and reliable articulatory information from the different vocal tract articulators involved in speech production (Narayanan, Alwan, & Haker, 1997) is quite challenging. Several increasingly accurate instrumental techniques of measuring vocal tract configurations have emerged in recent years. Amongst them, different imaging modalities (e.g., X-Ray, computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound (US)) have been used with the main goals of obtaining articulatory data, during speech production, and also for modeling purposes. However, the use of cineradiography and CT raised ethical concerns due to the use of ionizing radiation, precluding their use in speech production studies in healthy subjects. Ultrasound was used, e.g., to obtain tongue dorsum surface shapes (Stone & Lundberg, 1996), to evaluate lingual coarticulation (Zarkova, Hewlett, & Hardcastle, 2008) and has recently received particular at-

tention in Silent-Speech interfaces (Hueber et al., 2007). Though it provides an excellent temporal resolution, the information obtained is restricted to the dorsum of the tongue.

MRI is an imaging technique that uses strong magnetic fields, radiofrequency (RF) pulses and spatially varying magnetic fields to create high contrast images of biological tissues. In the last few years some important technological advances allowed important improvements not only related with the speed at which images can be acquired but also with the types of contrast that can be measured (Huettel, Song, & McCarthy, 2004). Contrary to the early days of MRI, when several minutes were required to create 2D images, nowadays it is common to use pulse sequences that acquire k-space data in three dimensions (volumetric imaging) in few seconds, or acquire images in real-time. Refinements include ultra-fast sequences (gradient and spin echo) high performance gradients, high field scanners, parallel imaging, new k-space sampling schemes (e.g., radial and spiral filling techniques) and extraordinary improvements in coil technology (Kim, Narayanan, & Nayak, 2009). Volumetric acquisition is especially desirable when high resolution anatomical images are needed. Exciting a whole volume of spins, instead of a slice, contributes to a high signal to noise ratio (SNR). Moreover, this yields thinner slices and better reconstructions in the orthogonal planes (Huettel et al., 2004).

In the field of speech production, MRI is one of the most promising techniques presenting some potential advantages: it provides good contrast between soft tissues (e.g., tongue, velum), allows 3D modeling (particularly relevant in the production of lateral sounds) and has the capacity of gathering information from the entire vocal tract (Narayanan et al., 1997). This last advantage is of special interest in the study of the pharyngeal cavity, since it is not accessible through other non-invasive articulatory techniques. Moreover, it is a non-invasive and safe imaging technique that does not involve the use of ionizing radiation. MRI limitations in the field are well known, including: difficulties in observing calcified structures (e.g., the teeth),

the acquisition technique, in which the speaker must be lying down during speech production (which may influence tongue posture) and the inherent noise produced during the acquisitions (expensive technology is required to record acoustic signal) (Engwall & Badin, 1999; Kitamura et al., 2005; Narayanan et al., 1997; Shadle & Damper, 2001). The relatively low temporal resolution achieved, even with the fastest acquisition techniques, is also a constraining factor if dynamic information is needed. However, some of these limitations have been partially overcome, with recent technological advances in the field, increasing the applicability of this imaging technique to speech.

For other languages (e.g., French, English, Swedish, German) several MRI studies (Badin, Bailly, Raybaudi, & Segebarth, 1998; Engwall & Badin, 1999; Narayanan et al., 1997) have been performed to obtain articulatory data or to model the vocal tract organs and the sounds it produces.

European Portuguese

For European Portuguese (EP), there is a lack of articulatory data, particularly, but not only, for specific classes of sounds such as laterals and rhotics (Oliveira, 2009). In the last few years a great investment in this field has been made (Martins, Carbone, Pinto, Silva, & Teixeira, 2008; Teixeira, Oliveira, Martins, Domingues, & Silva, 2009) with the main goal of developing an articulatory synthesizer, named SAPWindows (Teixeira et al., 2005). At this moment, oral vowels, nasal vowels and fricatives can be synthesized, with good results. Synthesis of other sound classes, such as laterals, will require 3D articulatory information (e.g., tongue shape, vocal tract areas, area and extension of the lateral channels) and dynamic data (e.g., nasal vowels, rhotics and diphthongs).

In this study, we use MRI to study articulatory characteristics of some EP sounds with the main focus on lateral consonants. EP has two lateral consonants: the /L/ (e.g., *palha*, *straw*) and /l/ (e.g., *sal*, *salt*). The /L/ is produced by one movement of the anterior tongue dorsum

against the alveolo-palatal area. The production of /l/ involves linguo-alveolar contact and the establishment of one or two lateral passages along the sides of the tongue. However, some interspeaker variability in the /l/ production has been reported and different patterns were found amongst different languages and dialects. In some languages, there are differences in the /l/ configuration depending on its position in the syllable: the /l/ produced at the end of a syllable has a more retracted tongue dorsum or elevation towards the velum than the /l/ produced at the beginning of the syllable. The latter is called *light-l* or non velarized /l/ and, the former a *dark-l* or velarized /l/ (Recasens & Espinosa, 2005).

For EP it is not clear if these differences in /l/ production occur or not. Previous acoustic (Andrade, 1999) and MRI data (Martins et al., 2008) seem to confirm the theory of the existence of a single configuration, in all syllabic positions. Moreover, it was not clear, for EP, how much the /l/ (and other sounds) is influenced when produced in the context of different vowels. The articulatory or acoustic influence of one speech segment on another is called coarticulation and is known to occur universally, but varies from one language to another (Farnetani & Recasens, 1999). Recent theories of speech production consider coarticulation crucial to models of speech production and speech synthesis (Shadle & Damper, 2001).

At present, our main goals concern the exploration of the 3D data, allowing an articulatory description of the sounds, and the extension of the results already obtained from 2D data. While, for some classes of sounds, 2D data (usually in the midsagittal plane) provides a rather good characterization, for lateral sounds (where air escapes over the sides of the tongue) 3D information is of the utmost importance. The purpose of the present study was to evaluate the feasibility of using a mesh comparison tool (PolyMeCo) in the study of different EP sounds; the same sound in different contexts or syllabic positions and comparisons amongst speakers.

In the next section the image acquisition technique is presented followed by a description

of the devised semi-automatic segmentation method. A short description of the tool used for comparison is also included. Relevant examples of the different comparisons performed are presented and discussed. Finally, the last section presents the main conclusions along with ideas for further work.

IMAGE ACQUISITION, SEGMENTATION AND POLYGONAL MESH COMPARISON

Figure 1 depicts the main steps involved in obtaining segmented data of the tongue: image acquisition, pre-processing, image segmentation and manual correction. In what follows, details concerning each of these steps are provided.

Image Acquisition

Seven EP speakers [three females (CO, MC, ER) and four males (JPM, JH, LCR, AS)], aged between 21 and 39 years, were recruited for the study. Six of the subjects are native monolingual EP speakers and one (JH) is a bilingual speaker (EP and Spanish). The speakers were all volunteers from the center of the country, with no history of hearing or speech disorders. An MRI screening form and informed consent were obtained before their participation in the study.

MRI data included the lateral-alveolar /l/ in different word positions: word initial (e.g., *laca*, *hairspray*), intervocalic (e.g., *sala*, *room*) and word final position (e.g., *sal*, *salt*) in the context of the three EP cardinal vowels ([i, a, u]). For the lateral /l/, only intervocalic context was considered (e.g., *palha*, *straw*).

The MRI data were collected at IBILI (Institute of Biomedical Research in Light and Image, Coimbra, Portugal), using a 3.0 T MR scanner (Magnetom Tim Trio, Siemens, Erlanger, Germany) equipped with high performance gradients ($G_{max} = 45 \text{ mT/m}$, rise time = 0.2s, slew rate = 200 T/m/s; and FOV = 50 cm). A standard 12-channel head and neck phased-array coils and parallel imaging (Generalized Autocalibrating Partially Parallel Acquisition

- GRAPPA) were used in all data acquisition sessions (refer to Figure 1, data acquisition step). The acquisition protocol was based on a previous MRI study (Martins et al., 2008). Subjects lay comfortably in a supine position in the MR machine using headphones.

After acquiring reference images, a T1 Weighted 5 mm thickness midsagittal MRI slice of the vocal tract was obtained using a Turbo Spin Echo (TSE) sequence (TR/TE/FA=400 ms/7.8 ms, 120°), FOV=240x240 mm; matrix (256 x 256) resulting in a pixel size of (0.938, 0.938). The acquisition time was 6 seconds. After that, a volume covering the entire vocal tract was obtained in the sagittal plane with a 3D ultra fast Spoiled GE sequence (Volume Interpolated Breath-Hold Examination - VIBE). This MR sequence can acquire T1-weighted images with high contrast and spatial resolution in a relatively short acquisition time (see Figure 1, data acquisition step). The sequence was originally applied to abdominal examinations but recently has also been used (Zheng et al., 2010) to study other structures (e.g., the knee). A comprehensive review of MRI pulse sequences can be found in (Bernstein, King, & Zhou, 2004). The volume comprises 52 slices, from the right to the left temporo-mandibular joint. The parameters used are TR = 4.32 ms, TE=1.37 ms, Flip Angle=10°, matrix (224 x 256), effective slice thickness = 2mm, voxel size (1.055, 1.055, 2) resulting in an acquisition time of 19 s. The speakers sustained the sound during the period of acquisition; the sequence was launched when the /l/ was produced.

Finally, a 3D high resolution sequence (VIBE) in the axial plane was obtained for each of the speakers, without phonation, to allow the extraction and co-registration of the mandible and maxilla masks. For each subject, the entire corpus was acquired in a single session lasting approximately one hour. Speech sounds from each subject were also recorded in an anechoic chamber during a separate session, in conditions that simulated the MRI environment (supine position and sustaining the target sound) to allow for the verification of the validity and consistency of the data.

Figure 1. Detailed pipeline used for data acquisition and segmentation. Data acquisition: MRI equipment used in the present study (left), 3D rendering of the volume obtained (middle) and midsagittal image (right). Pre-Processing (from left to right): Definition of the volume of interest (VOI), resampled region, and thresholded image. Segmentation (from left to right): 3D bubbles placed inside of the tongue, evolution after 121 iterations, and after 182 iterations when the evolution was finished. Manual correction (left to right): after examining the segmentation, some manual corrections must be performed.

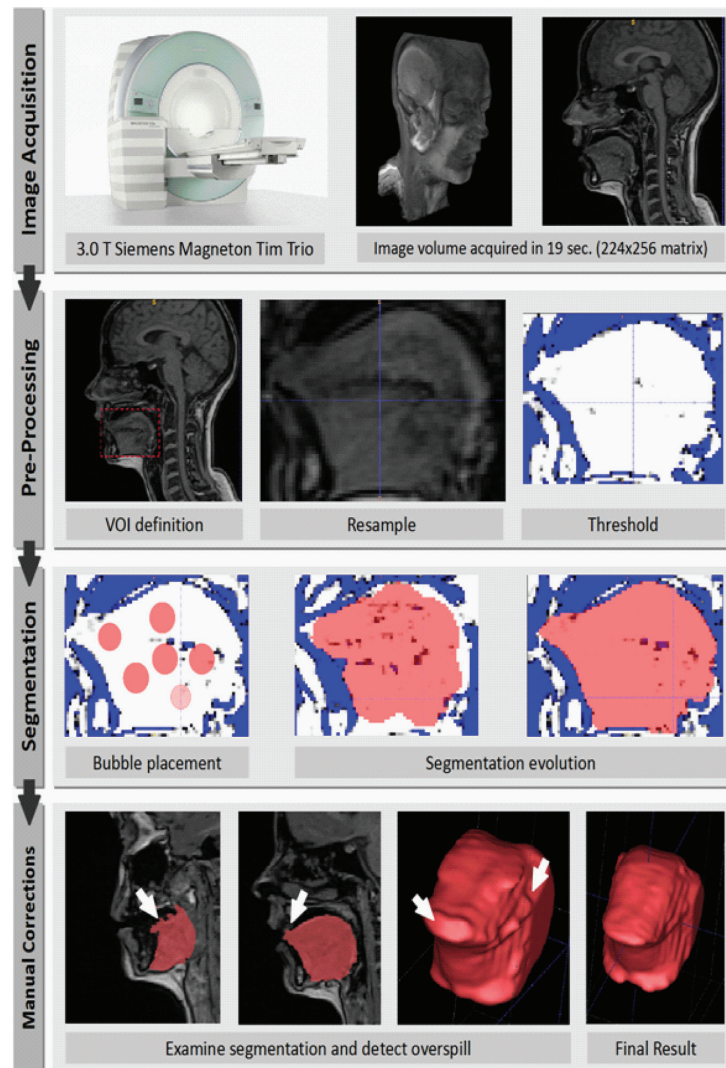


Image Segmentation

Segmentation, the process of partitioning images into constituent sub-regions that are homogeneous with respect to a specific characteristic, is a fundamental step in the formation of realistic tissue models (Pham, Chenyang, & Prince, 2000). Although other classifications are possible, segmentation approaches are usually divided into manual, semi-automatic and fully automatic techniques. Manual delineation of the object, performed by an expert, is frequently considered the gold standard, but this process is very time consuming, depends on the operator's prior knowledge regarding the structures to be segmented and, according to several authors (Pham et al., 2000; Poon, Hamarneh, & Abugharbieh, 2008; Yushkevich et al., 2006), reliability may vary. Fully automatic segmentation methods work best when their parameters are carefully tuned for specific image properties and anatomy (Poon et al., 2008). However, when the target anatomy is abnormal or pathology is present, which is frequent in clinical settings, reliability may suffer. Moreover, computational demands are very high. Semi-automatic segmentation techniques represent a good trade-off between efficiency and reliability and benefits from human expertise because some degree of user interaction and judgment is needed (e.g., initializing and choosing parameters) (Yushkevich et al., 2006).

In the field of speech production, segmentation of the structures involved in speech (e.g., tongue, velum, vocal tract) is a difficult and complex task, most often performed manually (e.g., Badin & Serrurier, 2006). For lateral sounds, these difficulties increase due to the formation of lateral side branches and of a separate supra-lingual cavity. Because we collected a rather extensive database (with seven speakers and several contexts), it is desirable to use semi-automatic tools. Segmentation of both the vocal tract and tongue images was based on the general deformable model framework. More details, on the segmentation of the vocal tract can be found in (Martins, Oliveira, Silva, & Teixeira, 2010).

In this article, our focus is tongue imaging. Tongue segmentation from MRI images was performed using ITK-SNAP toolkit (Yushkevich et al., 2006). ITK-SNAP has been used mainly in the field of neuroimaging to segment several intracranial structures from high resolution CT or MRI images (e.g., the caudate nucleus, lateral ventricles, hippocampus, brain tumors) with excellent reliability and efficiency (Ho, Bullitt, & Gerig, 2002; Ho, Cody, & Gerig, 2003; Yushkevich et al., 2006). Other examples include segmentation of the liver, bone structures for geometrical modeling (e.g., Lopes, Martins, Campos, & Pires, 2007; Ribeiro, Fernandes, Lopes, Folgado, & Fernandes, 2009), and the tongue from MRI images (Vogt, Wilson, Chander, & Fels, 2009).

This software is based on active contours or snakes. Snakes have become one of the most popular object delineation methods in image analysis particularly in biomedical and medical image analysis. Actually, all the snakes-based methods that have been developed may be understood as a special case of a more general framework for matching a deformable model to an image by means of some optimization procedure. The matching (or delineation, or evolution) process is described by a dynamical model where an initial curve or surface defined by the user, is iteratively pulled towards image-relevant features. Ever since the initial proposal described in Kass, Witkin, and Terzopoulos (1987), many approaches have been conceived to handle the model evolution taking into account criteria such as initialization tolerance, convergence speed, topology variance, external priors and noise sensitivity.

In this work, two sophisticated methods were evaluated under the implementation provided by the ITK-SNAP software tool: Geodesic Active Contours and Region Competition Snakes. Geometric deformable models were proposed independently by Caselles, Kimmel, and Sapiro (1997) and Malladi, Sethian, and Vemuri (1995) under the designation of Geodesic Active Contours or Level-set Front Propagation models respectively, as an alternative to the founding parametric active contour description.

Simpler parametric methods while suitable to many segmentation tasks are known to produce intersecting boundaries in many instances. To overcome this problem, the parametric description is substituted by a more general approach based on the estimation of curve evolution by solving numerically proper partial differential equations driven by only geometric constraints. Target curves or surfaces are implicitly embedded into higher dimensional functions and consequently evolution is determined by matching a particular level set into image-relevant features. This level set support may yield seamless segmentation of multiple objects, thus overcoming the problem of topological variance across the whole image dataset.

A tutorial overview of geometric deformable models may be found in Sonka, Hlavac, and Boyle (2007) while a more thorough treatment is given in Osher and Paragios (2003) or Osher and Sethian (1988). A more robust segmentation method within the snake realm was proposed by Zhu and Yuille (1996) and presented under the designation of region competition snakes. It is a rather elaborated approach where the merits of traditional snakes and statistical features of region growing are combined under a variational description where optimization is accomplished by a generalized Bayes/Minimum Description Length criterion.

In our study the region competition approach provided more accurate and faster results than the level-set approach due to 1) the specific context of our target object (the tongue); 2) the case of deploying suitable initialization seeds; and 3) the well-defined intensity difference between the tongue and the background image, i.e., air.

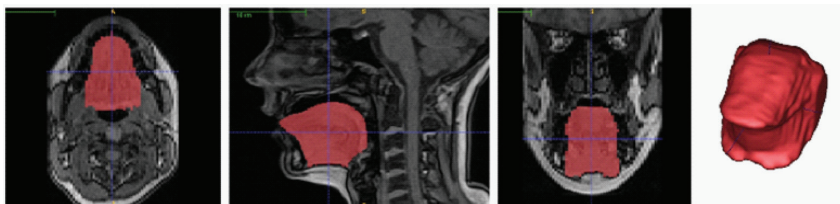
Segmentation Procedure

The steps to segment the tongue images semi-automatically include: 1) selecting the volume of interest (tongue), 2) resampling the volume to obtain an isotropic voxel, 3) defining a threshold, 4) placing seeds or bubbles inside of the tongue, 5) choosing the parameters that control contour evolution, and 6) finally run-

ning the algorithm (see Figure 1, pre-processing and segmentation steps). All of these steps take less than one minute. However, some time was spent previously in the optimization of all the parameters. After that, manual correction can be performed to obtain a more accurate segmentation (particularly at the lateral and inferior sides of the tongue), which can take approximately five additional minutes (revise the segmentation provided semi-automatically and remove or add unwanted or missing voxels. A mesh of the tongue (which can be smoothed and decimated) was obtained from the segmented volume using the well-known marching cubes algorithm (Lorensen & Cline, 1987) allowing a qualitative evaluation (e.g., position, overall tongue configuration, presence and location of grooving). Statistical data concerning the number of voxels and total volume is readily available from ITK-SNAP statistics. The mesh can be exported to other applications, using different formats (e.g., STL mesh format), offering the possibility of different analysis (e.g., comparing meshes in 3D) or for future modeling purposes, if desired. Figure 2 presents an example of the segmentation in the three orthogonal planes, and the resulting mesh.

In our study we devised an efficient technique for tongue segmentation, providing valuable 3D data. Nevertheless, this method needed validation in terms of reliability and accuracy. Accuracy is most difficult to evaluate when a real ground truth does not exist. Frequently, in these situations (e.g., clinical settings), manual tracings performed by an expert are compared with the results of other segmentation methods (Liu, Udupa, Odhnera, McDonough, & Arens, 2003). Therefore, for validation purposes fully manual segmentations were also performed using ITK-SNAP manual editing tool. Manual segmentation starts with delineation of landmarks in the axial and coronal images facilitating the identification of tongue margins, in the most lateral and inferior sides of the tongue. Afterwards the segmentation proceeds in the sagittal plane, slice by slice through the entire tongue. Finally a careful inspection and correction of the seg-

Figure 2. ITK-SNAP visualization of a MRI volume obtained for one of the female speakers CO (producing the /l/ as in the word sul, south. From left to right: views of the three orthogonal planes (axial, sagittal and coronal) and the mesh obtained after semi-automatic segmentation.



mentation must be performed. The process took an average of 30 minutes, ranging from 22 to 41 minutes depending on the speaker.

The segmentations performed semi-automatically were then compared with the manual segmentations using different quantitative measures including total volume comparisons and overlap measures such as Jaccard coefficient (Silva, Sousa Santos, Ferreira, Madeira, & Silva, 2011). The comparison results obtained pointed to an equivalent high reliability of both segmentation methods and a good agreement between the semi-automatic approach and the gold standard (manual) with the major advantage that the semi-automatic segmentation takes far less time to perform. Further details regarding the evaluation method and validation results can be found in Martins, Oliveira, Silva, Silva, and Teixeira (2011).

All the procedures described above were performed in a computer running Windows 7 with 6 GB RAM, 1.6 GHz Intel Core Duo processor and NVIDIA GeForce GPU.

Polygonal Mesh Comparison

Visual analysis and comparison of the polygonal meshes generated from the performed segmentations can serve two main purposes: 1) comparing meshes created from segmentations performed using the manual and semi-automatic methods allowed a simple qualitative assessment regarding how the latter method performed; and 2) an important improvement over visual analysis of the segmentations in 2D

planes. Nevertheless, comparing the meshes for different sounds (see Figure 3 for meshes corresponding to the cardinal vowels [i, a, u]), can still be a tiresome task, and no control is provided over the view conditions (e.g., view-point and zoom) to ensure they are the same for all compared meshes. If we aim to compare the same sound (e.g., the /l/), in different syllabic positions or contexts, this task is extremely difficult, because the sounds exhibit small differences (Figure 4).

There are several tools described in the literature such as Metro (Cignoni, Rocchini, & Scopigno, 1998), Meshdev (Roy, Foufou, & Truchetet, 2004) and PolyMeCo (Silva, Madeira, & Sousa Santos, 2009) which can be used for polygonal mesh comparison. PolyMeCo was chosen since it provides a user friendly environment where multiple mesh comparisons can be performed and analyzed side-by-side. This can be accomplished using several comparison data representations, such as histograms and colored meshes, while keeping synchronized view conditions (viewpoint, zoom, etc.) among all meshes. Figure 5 shows a snapshot of PolyMeCo's graphical user interface.

To analyze the comparison data shown in this article it is important to understand the basics of how this comparison is performed and how it is represented using a colored mesh.

When comparing two meshes one is considered the reference to which the other mesh is compared. For each vertex in the reference mesh a corresponding comparison point is found in the surface of the second mesh (which can be

Figure 3. Surface meshes of the tongue representing tongue configuration for the cardinal vowels [i] on the top, [a] on center, and [u] bottom, for the speaker CO. From left to right: anterior, lateral, posterior and oblique view.

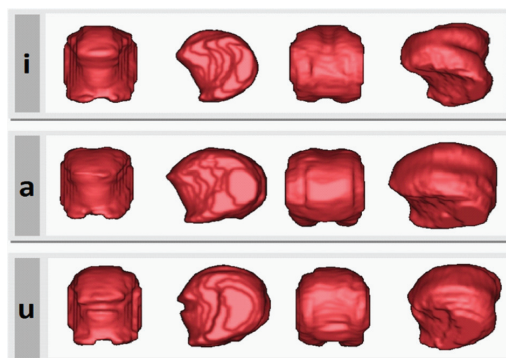
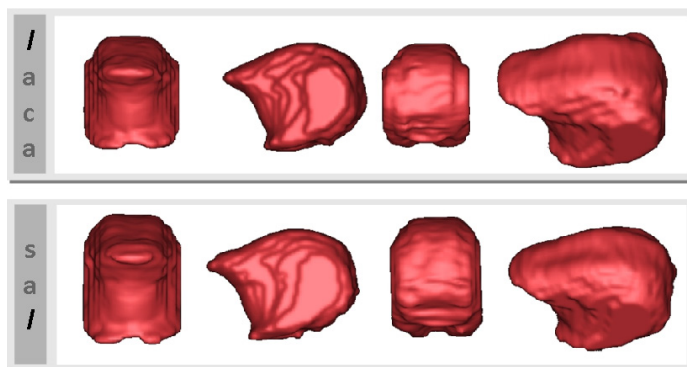


Figure 4. Surface meshes of the tongue representing tongue configuration for the lateral /l/ in different syllabic positions. Top, /l/ produced syllable initially, as in laca; Bottom: /l/ produced in syllable final position, as in sal. From left to right: anterior, lateral, posterior and oblique views.

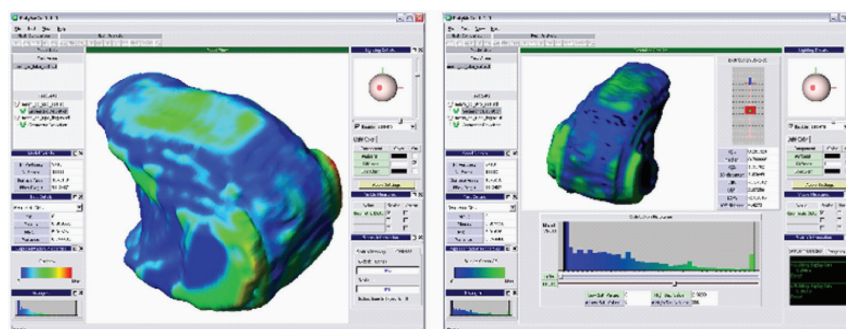


a vertex or a surface point, obtained by vertex interpolation), and a metric is used to compute their difference. PolyMeCo provides several metrics for comparing meshes using different criteria such as color, curvature and geometric distance. After computing the selected metric for all vertices in the reference mesh (towards their corresponding surface points in the second mesh) the obtained data can be represented by associating each computed value to the respec-

tive reference mesh vertex and presenting the reference mesh with each vertex colored accordingly. Therefore, the comparison data is always represented using the reference mesh.

For the comparisons performed in this article the geometric distance (GD) was considered the most adequate metric since it allows the assessment of differences between meshes, as intended, as can be easily interpreted.

Figure 5. PolyMeCo's graphical user interface. Left, default viewmode showing a mesh colored according to comparison data; right, viewmode showing a colored mesh and comparison data histogram.



RESULTS

A detailed articulatory description and a comprehensive phonetic interpretation of the findings are not provided as it is beyond the scope of this article. Instead, in this section, examples of the comparisons that can be performed are presented and discussed to highlight possible applications in the field of speech production (Figures 6 through 10).

Notice that in each figure all colored models use a common color mapping to allow comparison, i.e., the same color scale representing the same range of values.

The first example presented in Figure 6 allows comparison between the three EP cardinal vowels. It has been chosen because the differences between the cardinal vowels are more dramatic and prior knowledge concerning its articulatory description is available.

Briefly, concerning the position of the tongue in the oral cavity, the vowels are usually classified with respect to: 1) Tongue Dorsum (TD) height (high and low vowels, and 2) Tongue backness (anterior and posterior vowels). The [a] is classified (for EP) as a low back vowel, the [i] and [u] are high vowels; the latter is posterior and the former anterior. Accordingly, we expect differences between [i] and [a] at TD level due to differences in TD height,

and also at tongue root level due to differences in the degree of backness. In the comparison between [i] and [a] (see Figure 6, left), these differences are clearly demonstrated (in red at the back of the tongue, in the posterior view). Differences in TD height are observed in the oblique view (in yellow). Although, the differences are greater in the midline (probably due to a different degree of grooving) they spread through the sides. When comparing [a] and [u] (see Figure 6, right), differences are also detected in TD height and at the back of the tongue, although these are less prominent than those observed between [a] and [i].

Vowels and Consonants

Figure 7, on the left, presents comparison between the vowel [i] and the consonant /l/, in the context of [i] as produced in the word *litro*. Main differences between these two sounds occur at the tongue dorsum, tongue back and sides of the tongue. These differences can be easily justified: the gestures involved in the production of the vowel [i] and the /l/ are antagonistic, i.e., in the production of [i] the tongue is high and forward, while for the /l/ the tongue is retracted, somehow lowered in the middle part of the TD, and the tip is raised to establish contact with the alveolar area. Moreover, due to production requirements,

Figure 6. Mesh comparison between the three EP cardinal vowels as produced by the speaker (CO). The mesh corresponding to the vowel [a] was the reference mesh.

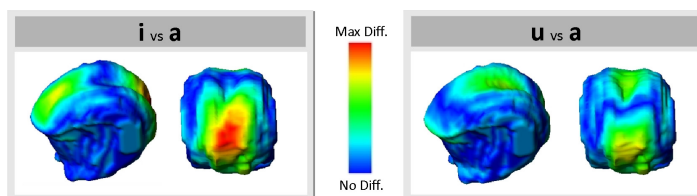
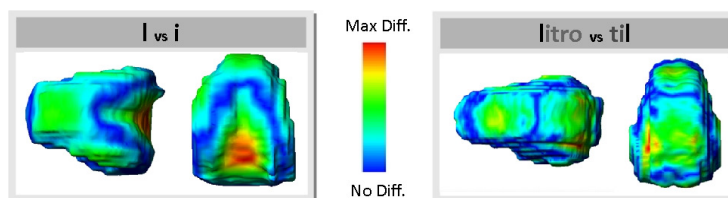


Figure 7. Mesh comparison between the vowel [i] (reference mesh) and lateral sound [l] when produced syllable initially followed by the vowel [i] (left). Mesh comparison between the /l/ in different syllabic positions as in *litro* and *til* (right).



in the /l/ the tongue is compressed towards the midline (accounting for the differences in the lateral sides of the tongue). In Figure 7, on the right, the /l/ in two word positions (*litro*, *til*) is compared. Differences observed on the back of the tongue can be attributed to a more retracted tongue in the production of the /l/ in syllable final position.

Much more complex comparisons can be performed, as presented in Figure 8 where the /l/, as produced in *laca*, is compared with the vowel [a], the lateral /L/ and the /l/ in different syllabic positions. It is interesting to note that there are fewer differences between [a] and /l/ with respect to differences between [i] and /l/ (refer to Figure 7), but the differences in the sides of the tongue are shared by both pairs.

Contextual Effects

The extent of lingual coarticulatory effects (i.e., the influence of the vowel in the production of the consonant) could also be qualitatively

evaluated by superimposition of the meshes representing the /l/ in the context of the three vowels as produced in the words *laca*, *litro* and *lupa*, and in the words *sal*, *til* and *sul*, as shown in Figure 9. As can be observed, the effect of the vowel in the production of the consonant is not prominent, and this effect is even more diminished when the /l/ is produced in syllable final position.

Inter-Speaker Variability

Furthermore, mesh comparison can also be used to compare and quantify differences between speakers (interspeaker variability) in the production of the same sounds. Figure 10 shows mesh superimposition of /l/ as in the words *laca* and *sal*, for each of the two speakers. As can be easily observed, differences between both /l/ productions are greater (extensive red color) for speaker JPM, than for speaker CO. Moreover, the information obtained from tongue comparison is in agreement with the pattern

Figure 8. Mesh comparison between /l/ (reference mesh), as produced in *laca*, compared with the vowel [a], the lateral /L/ and the /l/ in different syllabic positions

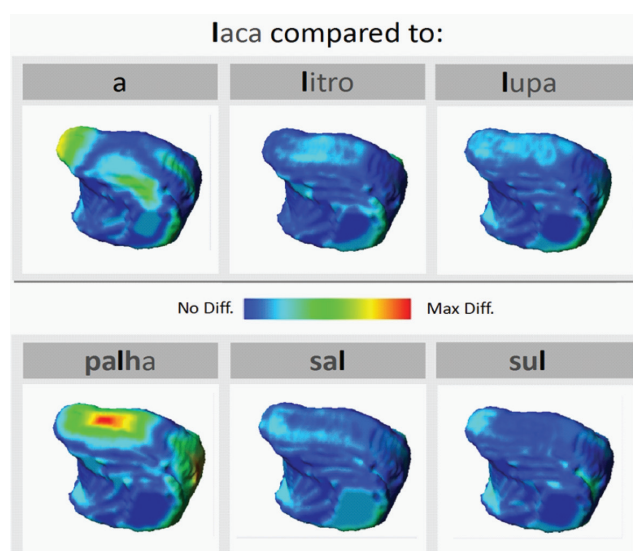
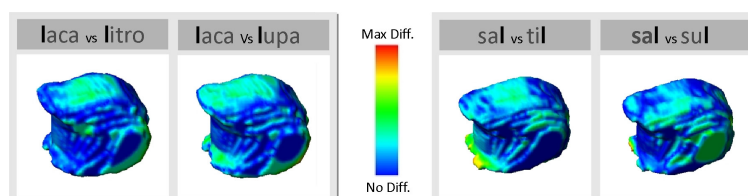


Figure 9. Multiple comparisons to evaluate contextual effects (coarticulation). Left to right: mesh comparison between /l/ in *laca* (reference mesh) with *litro* and *lupa* and mesh comparison between the /l/ in *sal* (reference mesh) with *til* and *sul*.



observed for vocal tract areas. The different analyses performed revealed that, as a rule, male speakers allowed more variability in their productions than female speakers. This is probably due to the different vocal tract dimensions between genders.

CONCLUSION

In this study, we used polygonal mesh comparison as a preliminary evaluation of tongue segmentation and to compare tongue surface meshes derived from MRI volumes acquired during the production of some EP sounds,

Figure 10. Comparison between speakers. Top (left): comparison of the /l/ as produced in *laca* and *sal* for the speaker JPM (male speaker), and at right corresponding area functions for the same sounds. Bottom: the same comparison for the speaker CO (female speaker); common color mapping was used.

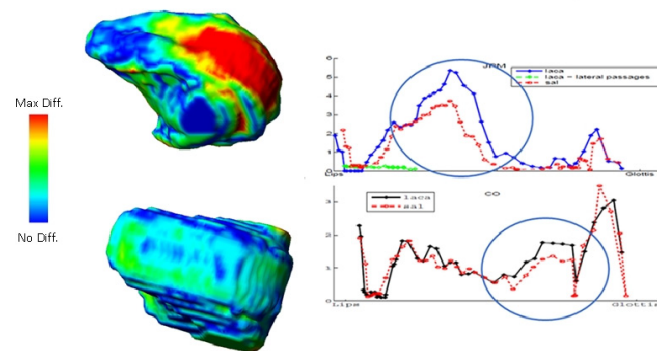
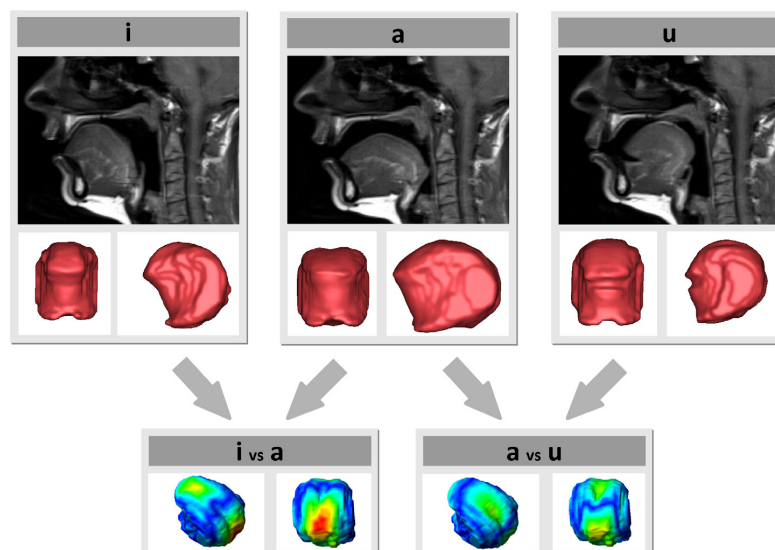


Figure 11. Sample illustrating how the MR images, meshes and mesh comparison can be used to teach articulatory phonetics



uttered by different speakers. Comparisons between different sounds, the same sound produced in different contexts and syllabic positions, and comparisons between speakers have been presented to illustrate the potentialities of this tool in the field of speech production. To the best of our knowledge no other study attempted this approach in order to compare 3D tongue shapes.

This is a qualitative comparison and the main interest, at the moment, is on where the differences occur rather than their quantification. Nevertheless, this comparison already provides very provocative results that provide a better understanding of the articulatory characteristics of EP speech sounds. In fact, the presented results are already useful, for example, to illustrate different aspects of speech production, for pedagogical purposes (Figure 11).

However, if coarticulatory and/or syllabic effects are to be further evaluated it will be necessary to quantify these effects. The values obtained for the geometric distance (GD), although representing a relative measure, are a first approach to this quantification. The metrics used for the validation process can also be used to provide a measure of coarticulatory effects or the effect of syllabic positions. To further explore gender, age or dialectal differences, a greater number of speakers and repetitions must be acquired. However, such an approach may require other instrumental techniques (e.g., EMA or ultrasound) due to the high costs and difficult access to MRI equipment.

At the moment we are working with normal speakers to model normal speech productions. In the future it will be very interesting to apply the computational tools explored in this article to subjects with speech disorders. Polygonal mesh comparison can be used to evaluate the extent of the deviation (e.g., tongue shape and posture) in disordered speech.

ACKNOWLEDGMENTS

This research was partially funded by FEDER through the Operational Program Competi-

tiveness factors – COMPETE and by National Funds through FCT (Foundation for Science and Technology) in the context of the Project HERON II (FCT reference PTDC/EEA-PLP/098298/2008). The first and second authors acknowledge FCT PhD grants respectively SFRH/BD/65183/2009 and SFRH/BD/38073/2007. The authors thank Professor Miguel Castelo-Branco (IBILI Director) and to the speakers involved in the study.

REFERENCES

- Andrade, A. (1999). On /l/ velarization in European Portuguese. In *Proceedings of the International Congress of Phonetic Sciences*, San Francisco, CA.
- Badin, P., Bailly, G., Raybaudi, M., & Segebarth, C. (1998). A three-dimensional linear articulatory model based on MRI data. In *Proceedings of the International Conference on Spoken Language Processing*, Sydney, Australia.
- Badin, P., & Serrurier, A. (2006). Three-dimensional linear modeling of tongue: articulatory data and models. In *Proceedings of the 7th International Seminar on Speech Production*, Ubatuba, Brazil.
- Bernstein, M., King, L., & Zhou, X. (2004). *Handbook of MRI pulse sequences*. Burlington, MA: Elsevier Academic.
- Birkholz, P., & Kröger, B. J. (2006). Vocal tract model adaptation using magnetic resonance imaging. In *Proceedings of the 7th International Seminar on Speech Production*, Ubatuba, Brazil.
- Caselles, V., Kimmel, R., & Sapiro, G. (1997). Geodesic active contours. *International Journal of Computer Vision*, 22, 61–79. doi:10.1023/A:1007979827043
- Cignoni, P., Rocchini, C., & Scopigno, R. (1998). Metro: measuring error on simplified surfaces. *Computer Graphics Forum*, 17(2), 167–174. doi:10.1111/1467-8659.00236
- Engwall, O., & Badin, P. (1999). Collecting and analysing two and three dimensional MRI data for Swedish. *TMH Quarterly Status and Progress Report*, 4, 1–28.
- Farnetani, E., & Recasens, D. (1999). Coarticulation models in recent speech production theories. In Hardcastle, W. J., & Hewlett, N. (Eds.), *Coarticulation: theory, data and techniques*. Cambridge, UK: Cambridge University Press.

42 International Journal of Creative Interfaces and Computer Graphics, 3(1), 28-44, January-June 2012

- Fels, S., Staveness, I., Hannam, A., Lloyd, J., Anderson, P., Batty, C., et al. (2009). Advanced tools for biomechanical modeling of the oral, pharyngeal, and laryngeal complex. In *Proceedings of the International Symposium on Biomechanics, Healthcare, and Information Science*.
- Gérard, J., Perrier, P., & Payan, Y. (Eds.). (2006). *3D Biomechanical tongue modeling to study speech production*. New York, NY: Psychology Press.
- Ho, S., Bullitt, E., & Gerig, G. (2002). *Level-set evolution with region competition: Automatic 3-D segmentation of brain tumors*. Paper presented at the International Conference on Pattern Recognition.
- Ho, S., Cody, H., & Gerig, G. (2003). *SNAP: A software package for user-guided geodesic snake segmentation (Tech. Rep.)*. Chapel Hill, NC: University of North Carolina.
- Hueber, T., Aversamo, G., Choller, G., Denby, B., Dreyfus, G., Oussar, Y., et al. (2007). *Eigentongues feature extraction from an ultrasound-based silent speech interface*. Paper presented at the International Conference on Acoustics, Speech, and Signal Processing.
- Huettel, S., Song, A., & McCarthy, G. (2004). *Functional magnetic resonance*. Sunderland, MA: Sinauer Associates.
- Kass, M., Witkin, A., & Terzopoulos, D. (1987). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), 321–331. doi:10.1007/BF00133570
- Kim, Y.-C., Narayanan, S., & Nayak, K. (2009). Accelerated three-dimensional upper airway MRI using Compressed Sensing. *Magnetic Resonance in Medicine*, 61, 1434–1440. doi:10.1002/mrm.21953
- Kitamura, T., Takemoto, H., Honda, K., Shimada, Y., Fujimoto, I., & Syakudo, Y. (2005). Difference in vocal tract shape between upright and supine postures: Observations by an open-type MR scanner. *Acoustical Science and Technology*, 26(5), 465–468. doi:10.1250/ast.26.465
- Liu, J., Udupa, J., & Odhnera, D., McDonough, & Arens, R. (2003). System for upper airway segmentation and measurement with MR imaging and fuzzy connectedness. *Academic Radiology*, 10, 13–24. doi:10.1016/S1076-6332(03)80783-3
- Lopes, D., Martins, J., Campos, J., & Pires, E. (2007). Modelação Geométrica de Estruturas Humanas baseada em imagens de Tomografia Computorizada. In *Proceedings of the CMNE/CILANCE*, Porto, Portugal.
- Loresen, W. E., & Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Computer Graphics*, 21(4), 163–169. doi:10.1145/37402.37422
- Malladi, R., Sethian, J. A., & Vemuri, B. C. (1995). Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17, 158–175. doi:10.1109/34.368173
- Martins, P., Carbone, I., Pinto, A., Silva, A., & Teixeira, A. (2008). European Portuguese MRI based speech production studies. *Speech Communication*, 50(11-12), 925–952. doi:10.1016/j.specom.2008.05.019
- Martins, P., Oliveira, C., Silva, A., & Teixeira, A. (2010). *Articulatory characteristics of European Portuguese Laterals: An 2D and 3D MRI Study*. Paper presented at the FALA VI Jornadas en Tecnología del Habla and II Iberian SLTech Workshop.
- Martins, P., Oliveira, C., Silva, S., Silva, A., & Teixeira, A. (2011). Tongue segmentations from MRI images using ITK-SNAP: Preliminary evaluation. In *Proceedings of the IADIS International Conferences on Computer Graphics, Visualization, Computer Vision and Image Processing*, Rome, Italy.
- Narayanan, S., Alwan, A., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals. *The Journal of the Acoustical Society of America*, 101(2), 1064–1077. doi:10.1121/1.418030
- Oliveira, C. (2009). *Do grafema ao gesto: Contributos para o desenvolvimento de um sistema de síntese de base articulatória* (Unpublished doctoral dissertation). University of Aveiro, Aveiro, Portugal.
- Osher, S., & Paragios, N. (Eds.). (2003). *Geometric level set methods in imaging, vision and graphics*. New York, NY: Springer.
- Osher, S., & Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi Formulation. *Computers in Physics*, 79(1), 12–49. doi:10.1016/0021-9991(88)90002-2
- Perrier, P., Payan, Y., Zandipour, M., & Perkell, J. (2003). Influences of tongue biomechanics on velar-stop consonants: A modeling study. *The Journal of the Acoustical Society of America*, 103, 1615–1627.
- Pham, D., Chenyang, X., & Prince, J. (2000). Current methods in medical imaging segmentation. *Annual Review of Biomedical Engineering*, 2, 315–337. doi:10.1146/annurev.bioeng.2.1.315

- Poon, M., Hamarneh, G., & Abugharbieh, R. (2008). Efficient interactive 3D Livewire segmentation of complex objects with arbitrary topology. *Computerized Medical Imaging and Graphics*, 32, 639–650. doi:10.1016/j.compmedimag.2008.07.004
- Recasens, D., & Espinosa, A. (2005). Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: evidence from two Catalan dialects. *Journal of the International Phonetic Association*, 35(1), 1–25. doi:10.1017/S0025100305001878
- Ribeiro, N., Fernandes, D., Lopes, D., Folgado, J., & Fernandes, P. (2009). *3-D solid and finite element modeling of biomechanical structures - a software pipeline*. Paper presented at the 7th EUROMECH Solid Mechanics Conference, Lisbon, Portugal.
- Roy, M., Fofou, S., & Truchetet, F. (2004). Mesh comparison using attribute deviation metric. *International Journal of Image and Graphics*, 4(1), 1–14. doi:10.1142/S0219467804001324
- Shadle, C. H., & Damper, R. I. (2001). *Prospects for articulatory synthesis: A position paper*. Paper presented at the 4th ISCA Workshop on Speech Synthesis, Seattle, WA.
- Silva, S., Madeira, J., & Sousa Santos, B. (2009). PolyMeCo - An integrated environment for polygonal mesh analysis and comparison. *Computers & Graphics*, 33, 181–191. doi:10.1016/j.cag.2008.09.014
- Silva, S., Sousa Santos, B., Ferreira, C., Madeira, J., & Silva, A. (2011). A preparatory study to choose similarity metrics for left-ventricle segmentations comparison. In *Proceedings of the SPIE Conference on Medical Imaging* (Vol. 7936).
- Sonka, M., Hlavac, V., & Boyle, R. (2007). *Image processing, analysis, and machine vision* (3rd ed.). Chonburi, Thailand: CL-Engineering.
- Stone, M., & Lundberg, A. J. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *The Journal of the Acoustical Society of America*, 99(6), 3728–3737. doi:10.1121/1.414969
- Teixeira, A., Martinez, R., Silva, L., Jesus, L., Principe, J., & Vaz, F. (2005). Simulation of human speech production applied to the study and synthesis of European Portuguese. *EURASIP Journal on Applied Signal Processing*, 9, 1435–1448. doi:10.1155/ASP.2005.1435
- Teixeira, A., Oliveira, C., & Barbosa, P. (2008). European Portuguese articulatory based text-to-speech: First results. In A. Teixeira, V. Lima, L. Oliveira, & P. Quaresma (Eds.), *Proceedings of the 8th International Conference on Computational Processing of the Portuguese Language* (LNCS 5190, pp. 101–111).
- Teixeira, A., Oliveira, C., Martins, P., Domingues, I., & Silva, A. (2009). *The project HERON*. Paper presented at the I Iberian SLTech - I Joint SIG-IL/Microsoft Workshop on Speech and Language Technologies for Iberian Languages.
- Vogt, F., Wilson, C., Chander, R., & Fels, S. (2009, June). Extraction of tongue shapes from magnetic resonance Images. In *Proceedings of the First International Workshop on Dynamic Modeling of the Oral, Pharyngeal and Laryngeal Complex for Biomedical Applications*, Vancouver, BC, Canada.
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., & Gee, J. C. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31(3), 1116–1128. doi:10.1016/j.neuroimage.2006.01.015
- Zarkova, N., Hewlett, N., & Hardcastle, W. (2008). *An ultrasound study of lingual coarticulation in children and adults*. Paper presented at the 8th International Seminar on Speech Production.
- Zheng, Z.-Z., Shan, H., & Li, X. (2010). Fat-suppressed 3D T1-weighted gradient-echo imaging of the cartilage with a volumetric interpolated breath-hold examination. *AJR. American Journal of Roentgenology*, 194, 414–419. doi:10.2214/AJR.09.2423
- Zhu, S., & Yuille, A. (1996). Region competition: Unifying snakes, Region growing, and Bayes/MDL for multi-band image segmentation. In *Proceedings of the Fifth International Conference on Computer Vision* (pp. 416–423).

44 International Journal of Creative Interfaces and Computer Graphics, 3(1), 28-44, January-June 2012

Paula Martins is an Assistant Professor at the Superior School of Health, University of Aveiro and a PhD student at the same university. She received a BSc in Radiography in 1987 and a MSc in Speech and Hearing Sciences in 2007, from University of Aveiro. She worked for 15 years as a Radiographer at Coimbra University Hospital. Her research interests include Speech Production, in particular Real time and Functional MRI applied to the study of Speech and Language. Additional interests include image processing and fine tuning of image acquisition protocols with an emphasis on image quality and patient radiation dose optimization. Paula Martins has been involved in different national research projects.

Samuel Silva is a PhD student at the University of Aveiro, Portugal, where he also received his degree in Electrical Engineering in 2003 and a MSc in Electrical Engineering in 2007. Since 2003 he has been working in the Institute of Electronics and Telematics Engineering of Aveiro (IETA). His research interests include geometric modeling, quality assessment of polygonal meshes, computer graphics, data visualization and medical imaging of the heart.

Catarina Oliveira is Assistant Professor at the University of Aveiro. She received her BSc degree in Languages and Literatures in 2003, from the University of Coimbra, and her PhD in Linguistics in 2009, from the University of Aveiro. Her current research interests include experimental phonetics, especially the articulatory and acoustic study of Portuguese sounds (e.g. nasal vowels), the impact of aging on voice and the applications of linguistic theory to speech synthesis. She has been involved in European and national funded research projects.

Carlos Ferreira is Invited Assistant Professor at the Superior School of Health, University of Aveiro and a Radiographer at the National Brain Imaging Network. He belongs to the investigation team of the Institute of Biomedical Research on Light and Image, Faculty of Medicine, University of Coimbra. His Research interests include functional MRI (fMRI) applied to the study of Speech and Language and real time RT-MRI.

Augusto Silva, PhD, is an Assistant Professor of Electrical Engineering at the Department of Electronics, Telecommunications and Informatics of the University of Aveiro. His research interests include Medical Imaging Systems, Medical Image Processing and Medical Imaging Informatics. He's been involved in several projects focused on image reconstruction, and image segmentation based on multi-dimensional datasets.

António Teixeira is Assistant Professor at University of Aveiro Electronics Telecommunications and Informatics Department. He received his BSc and Master in Electronics and Telecommunications Engineering in 1991 and 1993, and his PhD in Electronics Engineering in 2000, all from the University of Aveiro. He is Chair of the Special Interest Group on Iberian Languages of the International Speech Communication Association (ISCA) and the Director of the Master in Speech and Hearing Sciences. His main research activity is in the area of Speech Processing and Multimodal Human-Machine Interaction. He is involved in several national funded research (LUL, AAL4ALL, Smartphones for Smart Seniors) and in a starting European project (AAL Paelife), has supervised 4 concluded PhDs and more than 35 Masters, supervises more than 10 PhD projects and co-authored more than one hundred publications (books, book chapters, international journals and proceedings of international conferences).

Otimização de Protocolos e Procedimentos

Neste apêndice ao capítulo 6, pretendemos dar conta do processo de otimização conduzido no sentido de viabilizar as aquisições de imagem com RM em tempo real. Representa algum do trabalho preparatório efetuado durante as sessões experimentais. Os objetivos subjacentes a esta etapa foram:

- otimizar e selecionar as sequências de pulso a utilizar, em função dos objetivos;
- testar o microfone ótico e operacionalizar o processo de sincronização do sinal áudio com a aquisição de imagem;
- otimizar procedimentos para tornar mais rápidas e produtivas as sessões de recolha de dados e minimizar possíveis erros (e.g., codificação das imagens) que possam comprometer, à *posteriori*, a exploração dos dados.

F.1 Escolha e Otimização de sequências de pulso

Tendo em consideração que um dos objetivos delineados para esta parte do nosso trabalho é a obtenção de imagens de RM em tempo real, com uma resolução temporal que permita acompanhar a variação dos diferentes articuladores (e.g., língua, véu palatino, lábios) no tempo, uma das primeiras tarefas foi selecionar e identificar a sequência de pulso que melhor poderia servir os nossos objetivos e modificá-la até obter uma resolução temporal considerada adequada tendo em mente a informação pretendida (neste caso, o estudo de vogais nasais). O conhecimento adquirido durante o desenvolvimento do trabalho de Mestrado em Ciências da Fala (Martins, 2007), onde foi realizada uma experiência piloto com RM em tempo real¹, permitiu, desde logo, capitalizar algum conhecimento e balizar as nossas expetativas no que diz respeito à resolução temporal que se poderia vir a alcançar. Para além disso, permitiu-nos ter noção de alguns problemas que agora se tentaram ultrapassar.

Será útil elencar, desde já, alguns dos constrangimentos, de natureza variada, que condicionam a escolha e otimização deste protocolo, no âmbito da produção de fala. No geral, estes problemas são transversais a outros estudos.

¹No âmbito do trabalho de Mestrado, foi utilizada a sequência b-SSFP (TrueFISP), num campo de 1.5T, conseguindo-se obter 5 *frames/s*. Na altura não foi possível utilizar técnicas de imagem paralela. As imagens revelaram a presença de artefactos de suscetibilidade magnética, dificultando a visualização do véu palatino. Nalguns possíveis candidatos disponíveis para participar no estudo, os artefactos foram tão severos que nem se considerou a hipótese de adquirir essa parte do *corpus*.

Em RM há uma interdependência entre a resolução espacial, a resolução temporal e a S/R que obriga ao estabelecimento de alguns compromissos. De forma simples, para uma S/R aceitável, um aumento da resolução temporal pode ser conseguido à custa da degradação da resolução espacial. Por exemplo, a diminuição do número de passos de codificação da fase permite-nos tornar mais rápida a aquisição mas degrada a resolução espacial. O aumento da BW (permite diminuir o ESP) ou o recurso a técnicas de imagem paralela permitem melhorar a resolução temporal, no entanto, qualquer uma das abordagens contribui para diminuição da S/R se outros parâmetros não forem alterados.

No nosso estudo, é fundamental obter resolução temporal elevada² para acompanhar, no tempo, as variações dos articuladores na produção de alguns sons (e.g., vogais nasais). Todavia, é também desejável que se obtenha resolução espacial e S/R adequadas que facilitem a identificação das diferentes estruturas, permitindo, por exemplo, utilizar algoritmos de segmentação o mais automatizados possível. No que diz respeito ao processamento de imagem é sabido que quanto melhor for a resolução espacial e, menor o ruído, melhor o desempenho dos algoritmos. Como será necessário extrair os contornos dos vários articuladores estes parâmetros são críticos.

Outro grande problema, que é preciso evitar ou minimizar nas imagens de RM, são os artefactos. De pouco adiantará ter uma resolução temporal ou espacial “adequadas” se os artefactos na imagem condicionarem a sua interpretação. A quantidade de artefactos que podem surgir numa imagem de RM é imensa, com causas muito variadas (veja-se, por exemplo, uma revisão da literatura, bastante recente, relativa ao assunto apresentada por Morelli *et al.*, 2011). Frequentemente, a minimização de um artefacto pode originar outro se não houver cautela nas soluções que se tomam para a sua correção.

Neste estudo, são potencialmente preocupantes os artefactos de suscetibilidade magnética, de movimento, de *aliasing* (ver imagem F.1) e, outros, relacionados com aquisição rápida e com a utilização de esquemas de preenchimento ou subamostragem do espaço K (e.g., EPI, imagem paralela) que afetam em maior ou menor grau a qualidade de imagem. Os artefactos de suscetibilidade magnética, em particular, são uma grande fonte de preocupação, por várias razões:

- estes artefactos podem ocorrer nas interfaces ar/tecido devido à grande diferença de suscetibilidade magnética entre o ar (baixa) e o tecido (alta). Na zona do véu palatino, que neste estudo, vocacionado para o estudo das nasais, é o articulador mais importante, existe uma interface ar/tecido;
- para obter uma resolução temporal elevada, da ordem de várias *frames/s*, as sequências de pulso mais promissoras são, sem grandes dúvidas, as sequências EG, contudo, é bem conhecida a maior sensibilidade destas sequências, quando comparadas com as sequências SE ou TSE, a artefactos desta natureza (Elster e Burdette, 2001; Zhuo e Gullapalli, 2006; Morelli *et al.*, 2011). Este problema condiciona as nossas escolhas relativamente à sequência a utilizar;
- a intensidade de campo magnético do equipamento onde foram planeadas as aquisições é de 3T, o que também contribui para um aumento da sensibilidade às diferenças de suscetibilidade magnética entre os tecidos e a efeitos *off-resonance* em geral (Schick, 2005).

O processo de otimização das sequências decorreu num equipamento normal, para uso clínico, e não passou por efetuar alterações ao código fonte das sequências de pulso (e.g., alterar esquema aplicação de pulsos, ou dos gradientes) nem pela utilização de bobines dedicadas. A utilização destes recursos poderia permitir ir mais longe no que diz respeito à resolução temporal, no entanto, não foi previsto que assim acontecesse. As alterações efetuadas nas sequências de

²Note-se que não está claramente definido quais são os requisitos mínimos a cumprir em relação à resolução temporal ou espacial para estudar a fala. Como se referiu no capítulo 1, o ideal seria que as duas fossem elevadas. Não sendo possível, terá que se estabelecer o melhor compromisso entre estes dois atributos, que será diferente, por exemplo, em função dos sons a estudar ou do que se pretenda vir a caracterizar.

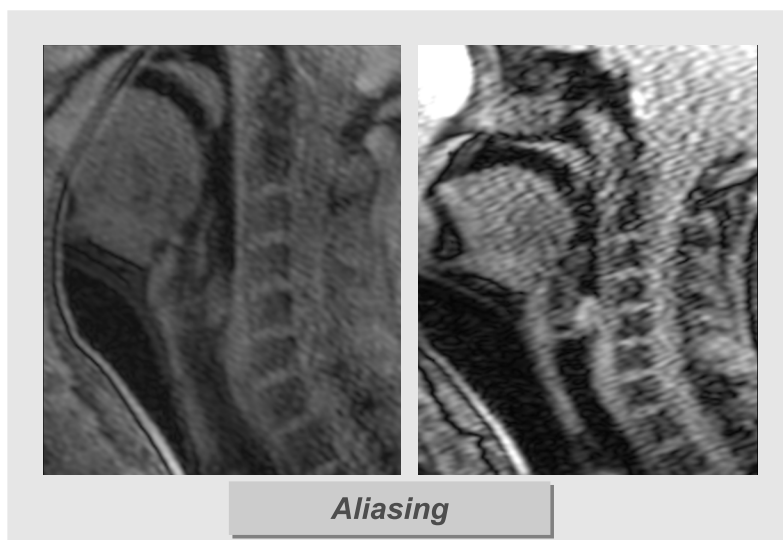


Figura F.1: Exemplo de artefactos de *aliasing* devido à escolha de FOV demasiado pequeno ao longo da direção de codificação de fase (sentido A-P).

pulso foram as que podem ser efetuadas na consola, pelo técnico de radiologia, com a manipulação de alguns parâmetros.

Apesar de existirem, no presente, sequências de pulso com preenchimento radial (e.g., FLASH radial) ou espiral e outras técnicas avançadas de aquisição rápida (veja-se o capítulo 2), que nos permitiriam uma resolução temporal mais elevada (e.g., *Compressing Sense*), o nosso leque de escolhas ficou condicionado aos pacotes de *software* disponíveis no equipamento onde os testes e aquisições de dados ocorreram.

Todas as sessões foram conduzidas dentro do modo normal de operação, ou modo controlado (também designado como *first level*) que é o que acontece, por rotina, quando se realizam estudos de RM. Desta forma, os valores de SAR e ENP estarão sempre dentro dos limiares considerados como seguros pelas diversas entidades reguladoras (FDA, 2003; ICNIRP, 2004; HPA, 2008). Sempre que possível optou-se por soluções que conduzissem a uma minimização do SAR e de ENP.

Apontados os constrangimentos, será também importante referir alguns fatores coadjuvantes, que auxiliaram o processo de otimização:

- Se existem algumas desvantagens associadas aos campos de intensidade elevada, também há claras vantagens, nomeadamente, o facto de a IS inerente ser mais elevada, com consequente melhoria da S/R. Isto permite-nos, no processo de otimização, alguma flexibilidade quando, para melhorar a resolução temporal, se tomam opções que conduzem a uma diminuição da S/R;
- Os gradientes que equipam o *scanner* são de elevado desempenho, apresentando valores de *rise time* e *slew rate* compatíveis com a obtenção de imagens com elevada resolução temporal;
- A existência de bobines *phased array* com número de canais adequados permitindo usufruir de técnicas de imagem paralela constituiu um importante contributo para a diminuição dos tempos de aquisição e para melhoria da resolução espacial. Para além disso, como já referimos no capítulo 2, a utilização de imagem paralela é extremamente importante nos campos elevados pois permite redução dos valores do SAR;
- Existem na instituição, dados os interesses de investigação subjacentes, alguns equipamentos e condições (físicas) que facilitam a utilização de dispositivos externos como o MO, que

noutras circunstâncias poderia ser muito mais problemático.

Tendo em atenção os aspetos referidos, nas primeiras sessões de testes, a tarefa principal foi testar (em fantoma e, posteriormente, em voluntários) sequências que nos permitissem obter, pelo menos, resolução temporal da ordem dos 100 *ms*. É evidente, como já referido, que as sequências que à partida melhor se posicionariam para obter razões de amostragem tão elevada seriam sequências ultra rápidas, TSE ou EG possivelmente conjugadas modo de aquisição EPI, com recurso a técnicas de subamostragem do espaço K e imagem paralela.

Desta forma, começou-se por procurar sequências dentro de protocolos já existentes, tais como protocolos para estudo cardíaco, monitorização da chegada de produto de contraste em estudos angiográficos ou para aquisição de imagem funcional. Refira-se que o ponto de partida para o processo de otimização foram sequências, com conjugação de parâmetros, em que o tempo de aquisição era da ordem dos segundos. Estes valores estão muito acima do que são os nossos objetivos no que diz respeito à resolução temporal.

Foram testadas sequências rápidas como EG -*Spoiled* (e.g., sequências FLASH e TurboFLASH), sequência SSFP (e.g., sequência FISP), EG *balanced Steady State* (TrueFISP) e, ainda, SS - TSE (HASTE) e EG-EPI. As figuras F.2 e F.3 apresentam exemplos de imagens obtidas com estas sequências, em fantoma e em voluntários.

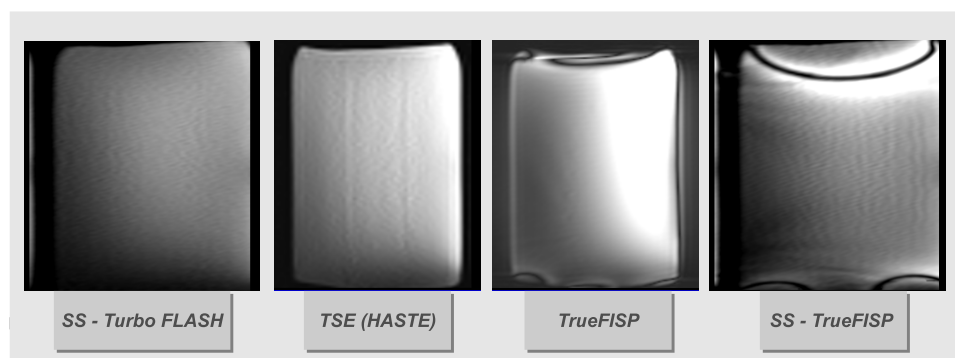


Figura F.2: Exemplos de algumas imagens (adquiridas com diferente número de *frames/s*), obtidas em fantoma, com sequências SS-TurboFLASH, SS-TSE (HASTE), TrueFISP e SS-TrueFISP. Artefactos de suscetibilidade magnética (*Banding artefacts*) mais evidentes com a sequência TrueFISP.

Os resultados das várias experiências efetuadas revelaram que a sequência com melhor potencial para nos permitir alcançar os objetivos propostos foi a sequência TurboFLASH. Foi com esta sequência que mais facilmente nos aproximamos da resolução temporal de 100 *ms* e foi a que apresentou, na generalidade, menos artefactos permitindo observar de forma adequada articuladores como a língua, o véu palatino e os lábios.

Um dos problemas com que nos deparamos diz respeito à S/R baixa (característica desta sequência) e com a perda de sinal, devido a efeitos de saturação, ao longo da sequência de imagens no tempo. Esta perda de sinal é, no entanto, um problema transversal a outras sequências de pulso como se pode observar na figura F.4. Em cima, imagens obtidas durante a fase de testes, com a sequência FISP e em baixo com a sequência TurboFLASH. Com a sequência FISP, para além da notória perda de sinal, a qualidade de imagem é bastante degradada pela presença de artefactos. A imagem da esquerda (para a sequência TurboFLASH), foi obtida logo no início da sequência ($t=144$ *ms*) e a da direita ao fim de 720 *ms* (após 10 imagens). É nesta fase inicial que a perda de sinal é mais acentuada. Note-se que, nesta série, a aquisição de imagem foi efetuada de forma contínua durante 5.6 *s* (75 imagens).

A sequência TrueFISP, apesar de ter uma grande potencial no que diz respeito à rapidez de aquisição e proporcionar uma IS elevada, apresentou, tal como já antecipávamos, muitos arte-

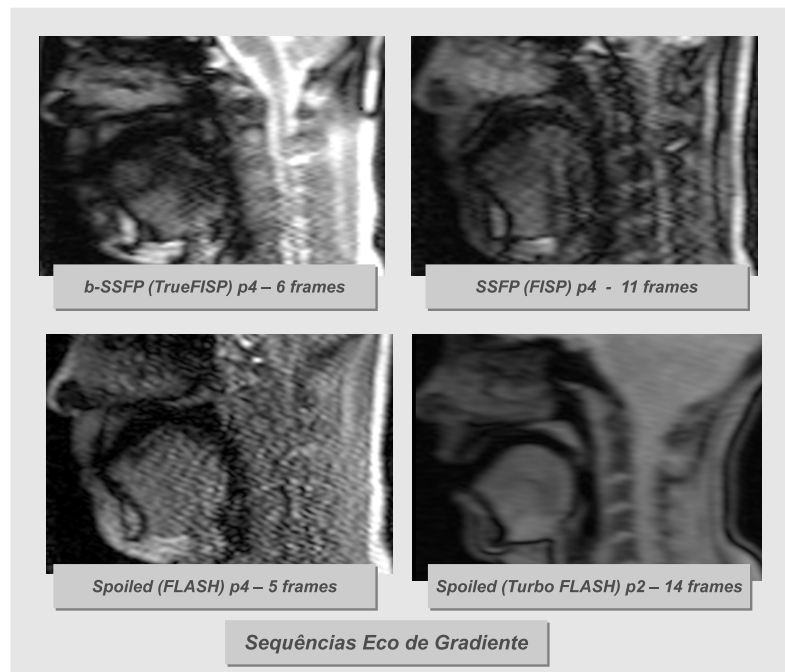


Figura F.3: Exemplos de imagens obtidas com algumas seqüências EG, testadas no âmbito deste estudo, com diferente resolução temporal (o número de frames indicado foi obtido num segundo).

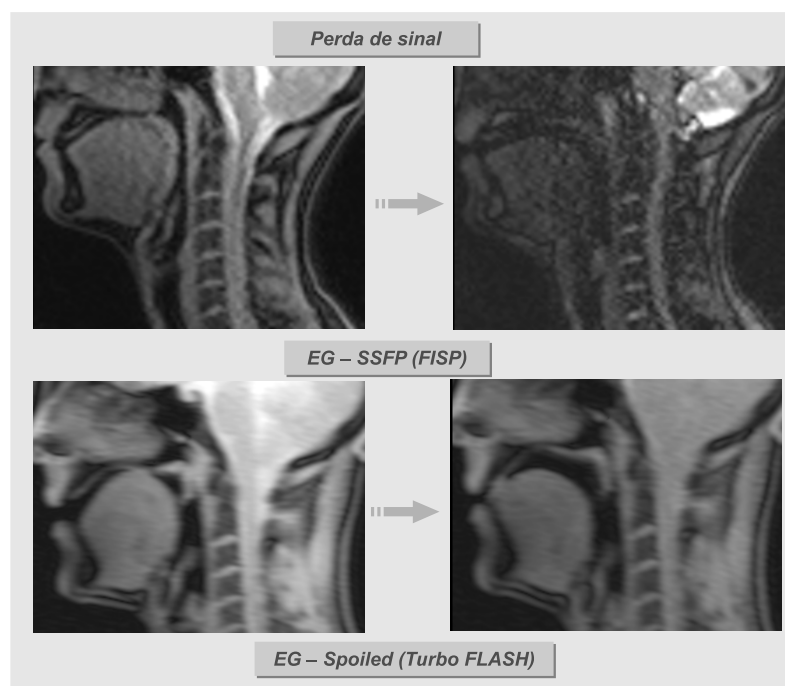


Figura F.4: Perda de sinal observada ao longo da seqüência de imagens devido a efeitos de saturação. Em cima, seqüência FISP e em baixo, a seqüência utilizada - TurboFLASH.

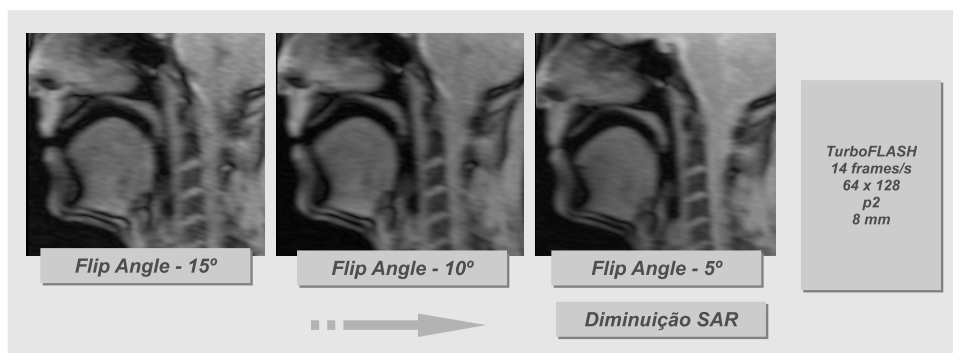


Figura F.5: Imagens adquiridas com a sequência TurboFLASH - Efeito da variação do *Flip Angle* e impacto no SAR.

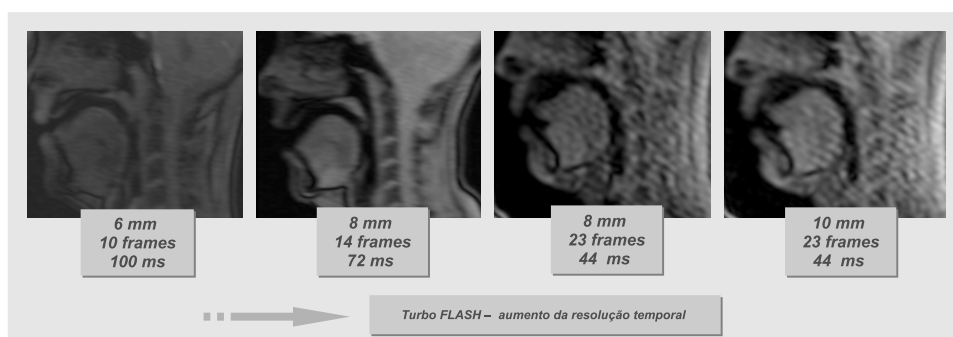


Figura F.6: Exemplo de imagens obtidas com a sequência otimizada, com diferente resolução temporal.

factos de suscetibilidade magnética na zona do véu palatino e da língua pelo que foi colocada de lado. Como se pode observar na figura F.2 obtida em fantoma e na figura F.3, já com voluntários, é esta a sequência que apresenta mais artefactos. Eventualmente, a utilização de técnicas de *shimming* localizado ou outras técnicas de correção poderiam minimizar estes artefactos. No entanto, a presença de qualquer material dentário (e.g., amálgamas) que a maior parte dos indivíduos apresenta poderia comprometer e dificultar a escolha dos falantes para o estudo. Uma sequência EG com modo EPI e a sequência HASTE foram também testadas mas as imagens obtidas, com ambas, foram de muito má qualidade. Neste caso, os vários artefactos que surgiram nas imagens (e.g., distorção) foram tão severos que impediram visualização das estruturas a estudar.

Assim, a opção mais viável foi trabalhar sobre a sequência TurboFLASH estabelecendo alguns compromissos para compensar a S/R baixa, nomeadamente um aumento da espessura de corte (de 6 mm para 8 mm), aumento ligeiro do FOV, diminuição do FA, etc. Por outro lado, o número de passos de codificação da fase foi diminuído, permitindo encurtar o tempo de aquisição à custa da degradação da resolução espacial; a BW foi aumentada para permitir ESP pequeno e diminuição do TE.

Mantendo todos os outros parâmetros constantes foram adquiridas imagens sem recurso a imagem paralela e com fatores de aceleração de 2 ($R = 2$) para verificar o ganho de tempo e as consequências na S/R e qualidade da imagem. Desta forma, conseguiu-se com a sequência TurboFLASH resolução temporal de 11 frames (96 ms) sem fator de aceleração e 14 frames/s (com $R = 2$). Foi com esta taxa de amostragem que se viria a adquirir o *corpus* de sons nasais uma vez que representou o melhor compromisso entre a resolução temporal requerida e a qualidade global das imagens. Na fase final, otimizou-se o *Flip Angle* tendo em mente uma menor perda de sinal e uma redução do SAR. Na figura F.5 apresentam-se imagens adquiridas com diferente FA. A qualidade da imagem não é, visualmente, muito diferente quando se passou de um FA de 10°

para 5°. No entanto, houve uma diminuição muito significativa do SAR (para cerca de $1/4$)³.

Posteriormente, com alguns ajustes a nível da matriz de aquisição e com fator de aceleração de 4 conseguimos uma resolução temporal de 23 *frames/s*. A figura F.6 apresenta exemplos de imagens adquiridas com diferente resolução temporal: à esquerda 100 *ms* (10 *frames/s*) com uma espessura de corte de 6 *mm* - é notório o baixo sinal mas melhor resolução espacial. De seguida, o protocolo utilizado nas aquisições (com 8 *mm* e 72 *ms*) - o melhor compromisso encontrado, tendo em atenção os objetivos do estudo. Ou seja, resolução temporal adequada para acompanhar os movimentos do véu palatino e dos outros articuladores mas, com qualidade que permitisse, por exemplo, extrair os contornos desses mesmos articuladores.

À direita, as imagens com resolução temporal mais elevada (44 *ms*) mas com óbvia degradação da resolução espacial e com mais artefactos. Ainda assim, contra o que eram as nossas expectativas (depois das alterações efetuadas), as imagens apresentam qualidade que nos parece suficiente para efetuar o seu processamento. Por exemplo, para o estudo dos sons róticos, onde esta melhoria da resolução temporal seria uma vantagem não será absolutamente necessário extrair contornos da língua mas será relevante observar e conhecer a duração do(s) batimento(s) ou vibração da língua na região alveolar (ou uvular).

F.1.1 Algumas considerações acerca da sequência TurboFLASH

Como foi descrito no capítulo 2, a sequência TurboFLASH é habitualmente utilizada com aplicação de pulsos para preparação da magnetização, podendo estes ser pulsos de saturação (*Saturation Recovery*) ou de inversão (*Inversion Recovery*). O objetivo é torná-la clinicamente mais útil, melhorando o contraste e a ponderação em T1 ou aumentar a S/R. O tempo desses pulsos preparatórios é variável e otimizável em função do que se pretenda avaliar ou caracterizar. Contudo, estes pulsos, apesar de só serem aplicados uma vez no início da aquisição (parte preparatória da sequência), condicionam a resolução temporal. Como neste estudo o objetivo é adquirir o maior número de imagens por segundo e não caracterizar lesões ou tecidos optou-se por utilizar a sequência sem aplicação de pulsos preparatórios, uma vez que conduziria ao aumento do tempo de aquisição.

A manipulação do FA nas sequências EG tem um papel importante no contraste e na ponderação das imagens obtidas. Neste contexto, não foi um motivo para grande preocupação uma vez que o que é relevante é a visualização das estruturas desejadas independentemente da imagem apresentar maior ou menor ponderação em T1 ou em T2. Assim, a utilização de um FA pequeno refletiu mais a necessidade de adquirir mais rápido, ter menos perda de sinal e diminuir o SAR do que propriamente para modelação do tipo de contraste proporcionado pelas imagens.

Relativamente à sequência utilizada houve, no entanto, algumas dúvidas relativamente ao valor do TR indicado (72 *ms*) e ao número de *frames* obtidas. Usualmente, o TR para aquisição de imagem em tempo real é muito mais pequeno do que aquele que se utiliza neste protocolo (e.g., abaixo dos 10 *ms*) levantando, na fase inicial, algumas questões que procurámos esclarecer e que tentaremos expor de forma sucinta.

Com um valor de TR tão elevado (72 *ms*) parecia impossível, por aplicação da fórmula apresentada no início do capítulo 6 ($R = \frac{1}{N_{Shot} \times TR \times NEX}$), obter 14 *frames/s*. Relembre-se que N_{Shot} é o número de pulsos de RF necessários para preencher todo o espaço K, NEX é o número de excitações. Numa aquisição 2D convencional o N_{Shot} será igual ao número de passos de codificação de fase. Desta forma, considerando que o TR é de 72 *ms* (como indicado nos parâmetros da sequência)⁴, com 48 passos codificação da fase ($64 \times 6/8$) e um fator de aceleração de 2 (técnica GRAPPA com 24 linhas de referência), o número de *frames/s* seria apenas de 0.4.

³Note-se que o cálculo efetivo do SAR é bastante complexo e exige dispositivos de medida adequados. No âmbito deste trabalho, limitamo-nos a comparar os valores de SAR fornecidos pelo equipamento e que constam num dos *tags DICOM*. O equipamento apresenta uma estimativa do SAR tendo em consideração o peso do doente, a sequência de pulsos utilizada, o número de cortes, etc.

⁴A totalidade dos parâmetros da sequência constam na secção 6.4.3.4 do capítulo 6.

Apercebemo-nos depois que o modo de aquisição da sequência utilizada é considerada pelo fabricante como um *Single-Shot* (SS). Recorde-se que, por definição, com uma técnica SS toda a matriz é codificada após um único pulso de excitação de RF (num TR). Desta forma, seria necessário apenas um pulso de excitação e o N_{Shot} seria 1. Assim, com $N_{Shot}=1$, NEX=1 e TR=72 ms obtemos as 14 frames/s o que estava de acordo com o que se tinha obtido e o problema ficou aparentemente solucionado.

No entanto, assumindo este raciocínio, não conseguimos explicar de que forma tinha sido efetuado o preenchimento do espaço K com esta sequência. Numa sequência SS-TSE ou HASTE com preenchimento cartesiano do espaço K é simples de explicar: após um único pulso de excitação (90°) seguido por uma série de pulsos de refocalização de (180°), tantos quantos o número de linhas da matriz (ou menos se for utilizada imagem paralela), toda a matriz é codificada. Com a utilização da técnica SS-EPI, de maneira diferente, após aplicação de um único pulso de excitação de RF a codificação do espaço K é efetuada, de forma extremamente rápida, através de oscilações bruscas no gradiente de leitura. A sequência TurboFLASH não parece encaixar em nenhuma destas duas alternativas para preencher o espaço K.

Dos diagramas de pulso encontrados na bibliografia, que representam a sequência TurboFLASH, nenhum nos permitiu esclarecer o assunto de forma satisfatória, uma vez que está sempre incluído o pulso preparatório que, no nosso caso, não se utilizou. Para além disso, apercebemo-nos que a própria definição do que é o TR, nesta sequência, não é muito consensual⁵ e que o próprio termo *Single-Shot* pode ser utilizado ou interpretado de forma diferente.

Das várias fontes consultadas, encontramos apenas uma referência que nos levou a perceber que a designação da sequência (tal como apresentada pelo fabricante) pode conduzir a erros de interpretação que estarão na origem da nossa questão inicial. De facto, Brix *et al.* (2008) referem que, por vezes, os fabricantes designam as sequências em que se utilizam pulsos preparatórios como sequências SS-TurboFLASH ou SS-TrueFISP mas, na realidade, estas não são verdadeiramente SS. Esta designação serve apenas para as diferenciar do modo de aquisição *Multi-Shot* ou de técnicas em que se utiliza espaço K segmentado.

“...rapid acquisition of a gradient-echo or steady state sequence following an inversion is sometimes referred to as single-shot technique. This is not quite correct since as many low angle excitation pulses are applied as Fourier lines are needed to fill the k-space. But the singleshot nomenclature allows a differentiation compared to the segmented, or multi-phase, imaging of the beating heart.” (Brix *et al.*, 2008, p. 72)

A oportunidade de efetuar uma simulação da sequência proporcionou um esclarecimento cabal do problema. A figura F.7 mostra a simulação da sequência TurboFLASH efetuada com recurso ao *software* IDEA-POET Versão VB17 (Siemens)⁶.

Com o POET foi possível simular a sequência escolhendo a categoria de sequências *EG-Spoiled* e introduzindo alguns parâmetros temporais (e.g., TR, TE, FA) e outros como a dimensão matriz, FOV, etc. Não é possível, no entanto, simular a utilização de imagem paralela o que no caso não foi problemático. Assim, foi simulado um dos protocolos que testamos, sem utilização de imagem paralela, com TR=96 ms TE=1.02 ms, matriz de 64×128 elementos com FOV parcial de 6/8 na direção da fase o que permite obter cerca de 11 frames/s.

Note-se que, para ser mais fácil a leitura do diagrama, na figura F.7, apenas se apresenta um *screen-shot* que corresponde a dois pulsos consecutivos de RF. Como é habitual num diagrama

⁵Por exemplo, a Siemens considera que o TR no TurboFLASH corresponde ao período compreendido entre a aplicação do pulso de inversão até ao preenchimento de toda a matriz de dados; Hashemi *et al.* (2012) consideram que o TR nesta sequência corresponde ao intervalo entre dois pulsos de RF, durante a fase de aquisição de dados.

⁶Agradecemos à Professora Doutora Sónia Isabel Gonçalves, investigadora no IBILI (Faculdade de Medicina da Universidade de Coimbra), a disponibilidade para efetuar a simulação da sequência uma vez que nos permitiu o esclarecimento das dúvidas levantadas relativas à sequência de pulsos utilizada. O ambiente IDEA ou *Integrated Development Environment for Applications* é uma ferramenta desenvolvida pela Siemens para desenvolvimento e teste de sequências de pulso para RM. A interface gráfica é designada como POET (*Protocol Offline Editing Tool*) que permite visualizar a sequência temporal e os vários eventos de uma sequência de pulsos.

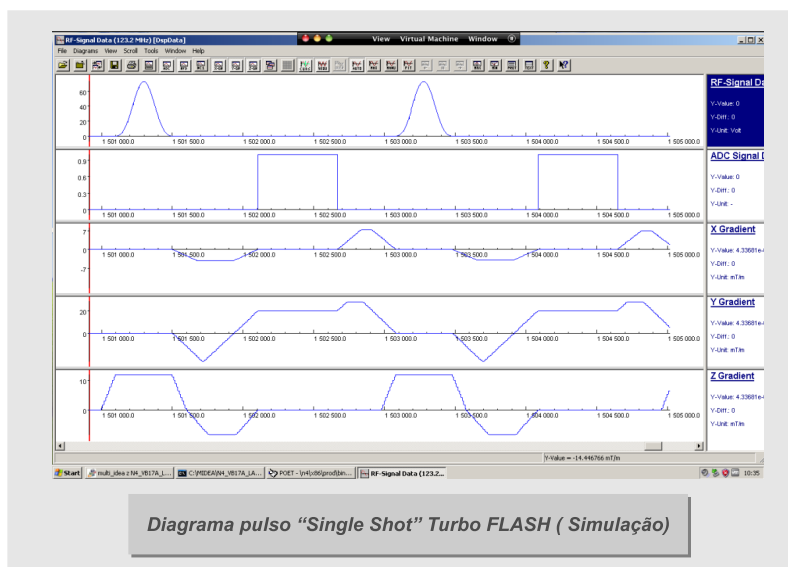


Figura F.7: Diagrama de pulsos da sequência TurboFLASH efetuada com o programa de simulação POET.

de pulsos estão representadas 4 ou 5 linhas que, neste caso, correspondem a: linha 1 - a modulação do pulso de excitação de RF; linha 2 - período de aquisição de dados (*ADC*) e, nas restantes linhas, o esquema de aplicação dos vários gradientes ao longo dos três eixos ortogonais (*X*, *Y* e *Z*). Assim, na linha 3, está representado o gradiente que efetua a codificação em fase, na linha 4 o gradiente de leitura e, na linha 5, o gradiente de seleção de corte.

Se observarmos a primeira linha é facilmente constataável que o intervalo entre dois pulsos de excitação de RF (usualmente designado como o *TR*) não é de 96 *ms* mas sim de 2 *ms*. Na realidade, verificamos que para formar uma imagem não é aplicado apenas um pulso de RF (como seria se fosse um verdadeiro SS) mas sim, neste caso, 48 pulsos de excitação (com *FA* pequeno - 5°) correspondentes ao número passos de codificação da fase. O que observámos está de acordo com o descrito por Brix *et al.* (2008) e por Hashemi *et al.* (2012).

Assim, especulamos que o “verdadeiro *TR*” será de aproximadamente 2 *ms* o que é bastante diferente do que nos é apresentado nos parâmetros da sequência uma vez que o fabricante considera como *TR* o intervalo de tempo necessário para obter uma imagem e designa-a como *Single-Shot*.

Voltando aos cálculos iniciais, considerando agora que o *TR* é de 2 *ms*, 1 NEX, 48 passos codificação da fase ($64 \times 6/8$), GRAPPA 2, com 24 linhas de referência (com este fator de aceleração apenas serão necessários 36 passos de codificação da fase para codificar a matriz) verificamos que o tempo de aquisição total para obter uma imagem é 72 *ms*, o que corresponde, efetivamente, a ter 14 *frames/s*.

Independentemente das questões terminológicas, que teriam relevância noutro contexto ou discussão, o importante foi verificar o *timing* a que foram aplicados os pulsos de RF e os gradientes de campo, permitindo clarificar a forma como se efetua a codificação da matriz com esta sequência.

F.2 Testes com Microfone ótico (MO), sincronização de imagem/som e codificação

Será oportuno referir que foi o nosso primeiro contacto com este tipo de tecnologia (MO) pois é a primeira vez que, de entre os estudos de RM que já efetuámos, foi realizada a gravação de sinal acústico de forma sincronizada com a obtenção de imagem. Nos estudos de produção de fala com aquisição de imagem em tempo real se o objetivo é obter informação temporal, torna-se absolutamente essencial efetuar a aquisição sincronizada (áudio e imagem), com tecnologia adequada, como se encontra bem documentado na literatura (Narayanan *et al.*, 2004; NessAiver *et al.*, 2006; Bresch *et al.*, 2006b). Relembre-se que o ruído provocado pela vibração das bobinas de gradiente durante as aquisições impede gravação do sinal acústico com microfones convencionais. Por outro lado, devido a questões de segurança, há restrições relacionadas com o material que pode entrar dentro da sala de RM, sendo necessário ter equipamento adequado e compatível com este ambiente.

Desta forma, tornou-se necessário, antes das sessões de aquisição, efetuar alguns testes com o MO para conhecer o seu funcionamento (e.g., processo de gravação, nível de cancelamento do ruído a selecionar), estudar a melhor forma de passar o cabo de fibra óptica para dentro da sala, uma vez que o dispositivo de controlo fica na sala da consola (fora da sala do magneto) e tem que ser conetado ao MO. Foi preciso também testar o *software* que o acompanha e perceber que tipo de ficheiros áudio gera. Numa segunda fase, houve necessidade de delinear uma forma que permitisse efetuar o alinhamento das imagens com o áudio.

O sistema FO-MRI-III utiliza dois microfones posicionados ortogonalmente, sincronizados em fase e amplitude, capturando o campo sonoro de entrada na mesma posição espacial. Utiliza algoritmos de redução de ruído que foram especificamente otimizados para utilização no ambiente de RM (gama dinâmica, frequência de ressonância etc). Os algoritmos comparam o sinal de entrada dos dois microfones para estimar com maior precisão o espectro de ruído e de voz ativa.

O *software* OptiMRI versão 3.1 (figuraF.8) permite, para além da gravação do *output* dos dois microfones, o processamento do sinal de fala filtrado e a possibilidade de ter até três pulsos *Transistor-Transistor Logic* (TTL), tudo sincronizado com elevada precisão. Cada gravação dá origem a uma série de ficheiros áudio referentes ao sinal captado pelos diferentes canais (*MAIN*, *REF*, *DSP-OUT*) e dos pulsos TTL.

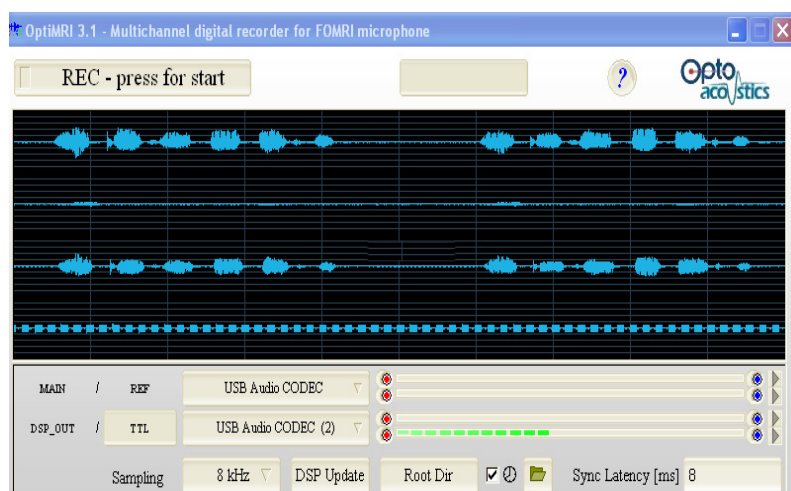


Figura F.8: *Output* do Microfone ótico.

Neste estudo foi utilizado apenas um pulso TTL, identificado com o nome *sinc*. O pulso TTL pode ser gerado através de um computador mas, a forma mais simples é que seja o próprio

sistema de RM (através da sequência de pulsos) a gerar o pulso. Desta forma foi necessário aceder ao código fonte da sequência de pulsos utilizada e programar a emissão de um pulso TTL a cada TR⁷.

Depois de definido o *corpus* a adquirir torna-se necessário efetuar a codificação de cada um dos itens de forma a poder identificar o falante, som, palavra, ou frase adquirida. Se tal não for efetuado, como a sequência de pulsos utilizada é sempre a mesma, não haveria forma de distinguir o que foi adquirido em cada produção. Esta codificação deve ser concebida de modo a fornecer o máximo de informação, deve ser clara simples e entendida por todos os intervenientes na aquisição. É este o código que numa fase posterior, durante o processamento de imagens, será utilizado para identificar, de forma rápida, uma imagem ou uma série de imagens numa base de dados com milhares de imagens. A utilização de símbolos fonéticos seria ideal uma vez que estamos a lidar com sons, no entanto, a sua utilização neste contexto deve ser evitada uma vez que pressupõe a utilização de símbolos específicos com que é difícil lidar nas consolas de aquisição de imagem.

No caso particular das aquisições em tempo real, para além das imagens existem ainda os ficheiro áudio que terão que ser associados, sem erros, a um conjunto de imagens. Foi durante as fases de testes com o MO que se tentou encontrar a melhor solução para identificar estes ficheiros. Por cada gravação, o MO gera e guarda, automaticamente, os ficheiros áudio resultantes de cada um dos canais e do pulso TTL, identificados pelo tempo de início da gravação do som. É possível renomear estes ficheiros, no entanto, o processo demora algum tempo e não é desejável perder tempo entre cada uma das séries do *corpus* sob pena de prolongar a sessão de aquisição de imagem. Assim, a forma mais simples e rápida de contornar a questão foi anotar o tempo fornecido pelo MO quando este inicia a gravação. Posteriormente, os registos áudio são renomeados com o nome do ficheiro de RM para a respetiva produção. De entre estes, os que vão ser utilizados são os referentes ao pulso TTL (*sinc*) e o DSP-OUT em que já foi efetuado o cancelamento do ruído.

Para tomar nota de toda esta informação é elaborada uma folha de registo ou um guião, para cada um dos falantes, que orienta as aquisições. Este registo prevê já a realização de sequências para localização e indica a ordem de aquisição de cada item. Este ordenamento não é aleatório, havendo a necessidade de estabelecer prioridades relacionadas com o que consideramos mais importante tendo em atenção os objetivos. Se por alguma razão um falante tiver que terminar a aquisição antes do previsto, o fundamental terá sido adquirido e a sessão não ficará totalmente comprometida. Na tabela F.1 apresentamos, a título ilustrativo, a informação constante da folha de registo com alguns itens do *corpus*.

Tabela F.1: Folha de registo das sessões de RM.

<i>Corpus</i>	N ^o sequência RM	Ficheiro RM	Tempo RM	Tempo som	N ^o imagens	Comentários
teste	2	cm-teste-sag	-	no sound	75	5.4 total
ampa, pampa, pan	7	cm-vnp-an-r1	17:35:00	17:51:57	75	-
infa, finfa, fin	25	cm-vnf-in-r2	17:50:44	18:07:46	75	-
pêpa	34	cm-voralp-e	17:59:27	18:16:25	75	repetiu 4 vezes
an, en, in, on, un	39	cm-vnasal-iso	18:03:31	18:20:14	300	repetiu 3 vezes
empa, pempa, pen	46	cm-obli-vnp-en-r1	18:19:20	18:36:15	75	-

Como se pode observar na tabela, este registo corresponde à falante CM, a codificação da série 7 indica-nos que foi adquirida a primeira repetição da vogal [6 ~] em contexto de [p]. A série 39, representa a série das vogais nasais isoladas e a série 46 corresponde à aquisição da vogal [e ~] em contexto de [p] no plano oblíquo. Por vezes, a informação pode parecer redundante no entanto, esta forma de agir tem-se revelado muito importante. De facto, do total de imagens adquiridas no âmbito deste estudo (para a falante CM temos 8067 ficheiros *DICOM*) nenhuma foi perdida e sempre que existiu alguma dúvida foi possível a identificação, da produção, de forma

⁷Agradecemos ao Dr. Gil Cunha, investigador do IBILI, a alteração do código fonte do TurboFLASH.

correta. O nome com que ficou guardada a série de RM consta de um dos *tags* DICOM e assim, para identificar este conjunto de imagens em MATLAB basta inserir o nome codificado.

Apêndice **G**

Corpus Nasais

G.1 Aquisições em Tempo Real - Nasais

G.1.1 Vogal nasal em posição inicial, medial e final - Contexto de oclusiva bilabial [p]

1. ampa, pampa, pan - [6~p6], [p6~p6], [p6~]
2. empa, pempa, pen - [e~p6], [pe~p6], [pe~]
3. impa, pimpa, pin - [i~p6], [pi~p6], [pi~]
4. ompa, pompa, pon - [o~p6], [po~p6], [po~]
5. umpa, pumpa, pun - [u~p6], [pu~p6], [pu~]

Três repetições de cada sequência

G.1.2 Vogal nasal em posição inicial, medial e final - Contexto de fricativa [f]

1. anfa, fanfa, fan - [6~f6], [f6~f6], [f6~]
2. infa, finfa, fin - [i~f6], [fi~f6], [fi~]
3. unfa, funfa, fun - [u~f6], [fu~f6], [fu~]

Três repetições de cada sequência

G.1.3 Vogais orais produzidas em contexto de oclusiva bilabial [p]

1. Pápa - [pap6]
2. Pâpa - [p6p6]
3. Pépa - [pEp6]
4. Pêpa - [pep6]
5. Pipa - [pip6]
6. Pópa - [pOp6]
7. Pôpa - [pop6]
8. Pupa - [pup6]

Repete a mesma palavra durante a aquisição

G.1.4 Sequência de vogais orais e nasais produzidas de forma isolada

1. an, en, in, on, un - [6~],[e~], [i~], [o~], [u~]
2. a é i ó u (vogais orais abertas) - [a, E, i, O, u]
3. â ê ô (vogais orais fechadas) - [6, e, o]

Repete a sequência durante a aquisição

G.1.5 Oblíquos e duplos oblíquos

1. ampa, pampa, pan
2. empa, pempa, pen
3. impa, pimpa, pin
4. ompa, pompa, pon
5. umpa, pumpa, pun
6. anfa, fanfa, fan
7. infa, finfa, fin
8. unfa, funfa, fun

Três repetições de cada sequência

G.1.6 Consoantes nasais - posição intervocálica

1. Ama, ana, anha - [6m6], [6n6], [6J6]
2. Imi, ini, inhi - [imi], [ini], [iJi]
3. Umu, unu, unhu - [umu], [unu], [uJu]

Duas repetições da mesma sequência

G.1.7 Aquisição com frase suporte - Diz XXX após

1. ampa, pampa, pan
2. empa, pempa, pen
3. impa, pimpa, pin
4. ompa, pompa, pon
5. umpa, pumpa, pun

Só uma aquisição por item

G.2 Material Complementar - Laterais e róticos

1. laca, sala, sal - [lak6, sal6, sal]
2. litro, bilis, til - [lit4u, biliS, til]
3. lupa, pulo, sul - [lup6, pulu, sul]
4. caro
5. carro

6. carro

7. Frase livre à escolha do falante

Repete a sequência ou item durante a aquisição